UNIVERSITY OF ALBERTA

EFFECTS OF TEST-WISENESS UPON PERFORMANCE ON THE TEST OF ENGLISH AS A FOREIGN LANGUAGE

BY

PING YANG 

A thesis submitted to the Faculty of Graduate Studies and Research

in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

Edmonton, Alberta

Fall, 2000

ABSTRACT

The major objective of this study was to investigate the influence of test-wiseness upon performance on the Test of English as a Foreign Language (TOEFL). In order to achieve this objective, methods involving both quantitative and qualitative approaches were employed. First a sample of 390 Chinese TOEFL candidates was selected and asked to respond to a modified version of the Test of Test-wiseness (TTW) and the TOEFL Practice Test B (ETS, 1995d). Based on the results of item analysis of the TTW, a subsample of 40 students, consisting of 23 test-wise and 17 test-naïve students, was identified and then interviewed. In the interview, each student was asked to "think aloud" about the strategies she/he was using while responding to the Interview Form containing some test-wiseness susceptible items selected from the TTW and TOEFL. Students' responses were recorded and analyzed.

The presence of test-wise susceptible items in the TOEFL was initially identified by the two judges' consensus judgements, then verified by empirical evidence obtained from the item analysis of the TOEFL based on the responses of 390 Chinese students. Items were not considered to be susceptible to test-wiseness unless identified by both the judgmental and empirical processes.

To further validate the findings pertinent to the presence of test-wise susceptible items in the TOEFL and to understand what general cognitive processes the Chinese students usually applied when responding to test-wise susceptible items in the TTW and TOEFL, protocol analysis (Ericsson & Simon, 1993) of the interview data was conducted. A comparison was made between test-wise and test-naïve students in terms of the strategies used when responding to the test-wise susceptible items in the TTW and

TOEFL.

Based on the evidence gathered from the judgmental and empirical analyses, 48% to 64% of the items across the Listening and Reading Comprehension subtests of the TOEFL Practice Test B (ETS, 1995d) were identified as susceptible to test-wiseness. According to the results of a correlated t-test, the mean scores for the items identified as susceptible to test-wiseness exceeded significantly ($p < 0.01$) the corresponding mean scores for the non-susceptible items in both the Listening and Reading Comprehension subtests.

A preliminary analysis of the TTW suggested that approximately 70% or more of the 390 Chinese students were able to identify and use the cues related to absurd options, similar options, and opposite options. The results of the protocol analysis of the interview data further revealed that the differences between the test-wise and test-naïve students were not only in the quantity but also in the quality and effectiveness of using the appropriate test-wise skills to figure out the answers not known by their content knowledge alone. Compared to the test-naïve students, the test-wise students' approaches appeared to be more meaningful, thoughtful, logical, and less random. In addition, they were more academically knowledgeable and more persistent in looking for test-wiseness cues than their test-naïve counterparts.

When responding to the selected TOEFL items, especially the test-wise susceptible items in the Interview Form, the test-wise students consistently outperformed the test-naïve students in terms of the proportions to get the correct answers. The results of the "think-aloud" protocol analysis showed that test-wise students' higher performance on the susceptible items may be explained by the "construct-irrelevant easiness"

(Messick, 1989) achieved through the combination of test-wise susceptible items, the possession of test-wiseness skills, and partial knowledge. If the ability to apply test-wiseness strategies combined with partial knowledge is considered irrelevant to the construct measured by the TOEFL, it is suggested that the current administration procedures and the format of the TOEFL need to be reviewed and improved.

ACKNOWLEDGEMENTS

First, I wish to express my profound appreciation and sincere thanks to Dr. Todd Rogers, my program and thesis advisor, for taking the challenge and the full responsibility to guide me through the program and thesis writing. His heart-warming encouragement and care, his unselfish and unconditional support, his invaluable expertise in measurement and evaluation, his quick response and insightful suggestions, and his rigorous scientific approach made it possible for me to complete this massive study. His excellent instruction in psychometrics was, and will continually be, the guidance in my professional career. My heartfelt gratitude toward him is far beyond any verbal expression.

Also, I would like to express my sincere thanks and special gratitude to the following people:

Dr. Thomas Maguire for his masterly instruction in psychometric theory and practice. His great expertise and extensive knowledge of measurement and evaluation have enlightened me and contributed substantially to my education.

Dr. Mark Gierl for his long-term support and guidance in the field of measurement and evaluation. His expertise in psychometrics, particularly in item response theory and equating, has made a unique contribution to my professional development.

Dr. Bruce Derwing for his invaluable expertise in linguistics and in the TOEFL preparation instruction which contributed substantially to this research.

Dr. Leonard L. Stewin, Dr. Gay L. Bisanz, Dr. Walter Muir, and Dr. Tracey Derwing for serving on my Examining Committee and providing insights into this study.

In addition, I would like to thank faculty members who have contributed to my professional development and thank the staff who provided me with the full assistance whenever I needed it.

I would like to express my heartfelt gratitude to my wife and to my son for their deep love, their understanding and tolerance and their full support over years. Without their support, the completion of this study could never have been possible.

My ultimate gratitude goes to my Canadian sponsor, Professor Lars Thompson at the Queen's University, for his extreme generosity and father-like love and care. Without his support, I could never have made the trip to Canada.

I would like to acknowledge the Educational Testing Service for providing me with the TOEFL test material and for allowing me to use the test material in this study.

I would like to acknowledge the president, the vice-president and their cooperating students at Shanghai Qianjin Institute in P. R. China for allowing me to collect data.

This research was financially supported in part by the Fund for the Support of International Development Activities (FSIDA), a Social Science and Humanities Research Council of Canada (SSHRC) Doctor Fellowship, a Walter H. Johns Graduate Fellowship, an Edward Chang Memorial Graduate Scholarship, and a University of Alberta PhD Scholarship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The Rationale for the Research

As the largest and the most influential English test in the world, the Test of English as a Foreign Language (TOEFL) is taken by approximately 800,000 nonnative speakers of English from 180 countries each year (Educational Testing Service, 1995b). It provides scores, obtained from multiple-choice items, that contribute to decisions regarding admission to or exclusion from more than 2,400 colleges and universities in the United States and Canada. Given the importance of the decisions made on the basis of the TOEFL scores, numerous studies on test validation, reliability, and utility have been carried out by the Educational Testing Service (ETS) as well as many other researchers (ETS, 1995a). To date, however, little research has been conducted to investigate the influence of test-wiseness upon performance on the TOEFL.

Test-wiseness is "a subject's [examinee's] capacity to utilize the characteristics and formats of the test and/or the test taking situation to receive a high score" (Millman, Bishop, & Ebel, 1965, p. 707). If an examinee possesses test-wiseness, and if the test contains susceptible items, then the combination of these two factors can result in an inflated test score; in contrast, a test-naïve examinee, i.e., an examinee with little test-wiseness, might likely be penalized whenever the test involves susceptible items (Rogers & Bateson, 1991a, b). Given the comparative nature of many of the uses of test results, an issue of fairness arises. Further, although it might be expected that major standardized tests, such as the TOEFL, developed by professionals through a series of rigorous

technical procedures (Peirce, 1992) would be relatively immune to test-wiseness, research suggests that this is not always the case (Rogers & Yang, 1996; Rogers & Bateson, 1991a, b; Rogers & Wilson, 1993; Benson, 1988; Bangert-Drowns, Fagley, 1987; Hughes, Salvia & Bott, 1991; Metfessel & Sax, 1985; Kulik, & Kulik, 1982; Smith, 1982; Slack & Porter, 1980; Sarnacki, 1979). For example, Hughes et al. (1991) found that approximately 75% of teacher-made and publisher-provided tests contained items susceptible to test-wiseness. Rogers and Bateson (1991a) found that across six Grade 12 provincial school-leaving examinations, the percentage of test-wise susceptible items varied between 43% and 80%. These findings, together with the consideration of the great influence of the TOEFL scores on the tertiary admission process and on professional certification in North America, justify the need to assess empirically the effects of test-wiseness on the TOEFL. If TOEFL scores are subject to the influence of test-wiseness, then people involved with test development, administration, and interpretation should be well informed of the construct of test-wiseness and how it may affect test scores.

Equally important and relevant to the issue addressed above is research on test-taking behaviour of test-wise examinees. Research on this topic will facilitate a better understanding of test-wiseness and shed light on why some of the items are susceptible to test-wiseness (Bachman, 1990). One of the viable instruments to conduct research of this kind is a model of test-taking behaviour proposed and verified by Rogers and Bateson (1991a, Figure 1). Based upon the work of Brown (1980; 1987), Flavell (1979), and Schuell (1986), and an early model proposed by Smith (1980), the model reflects various routes an examinee, especially a test-wise examinee, may take to determine which option

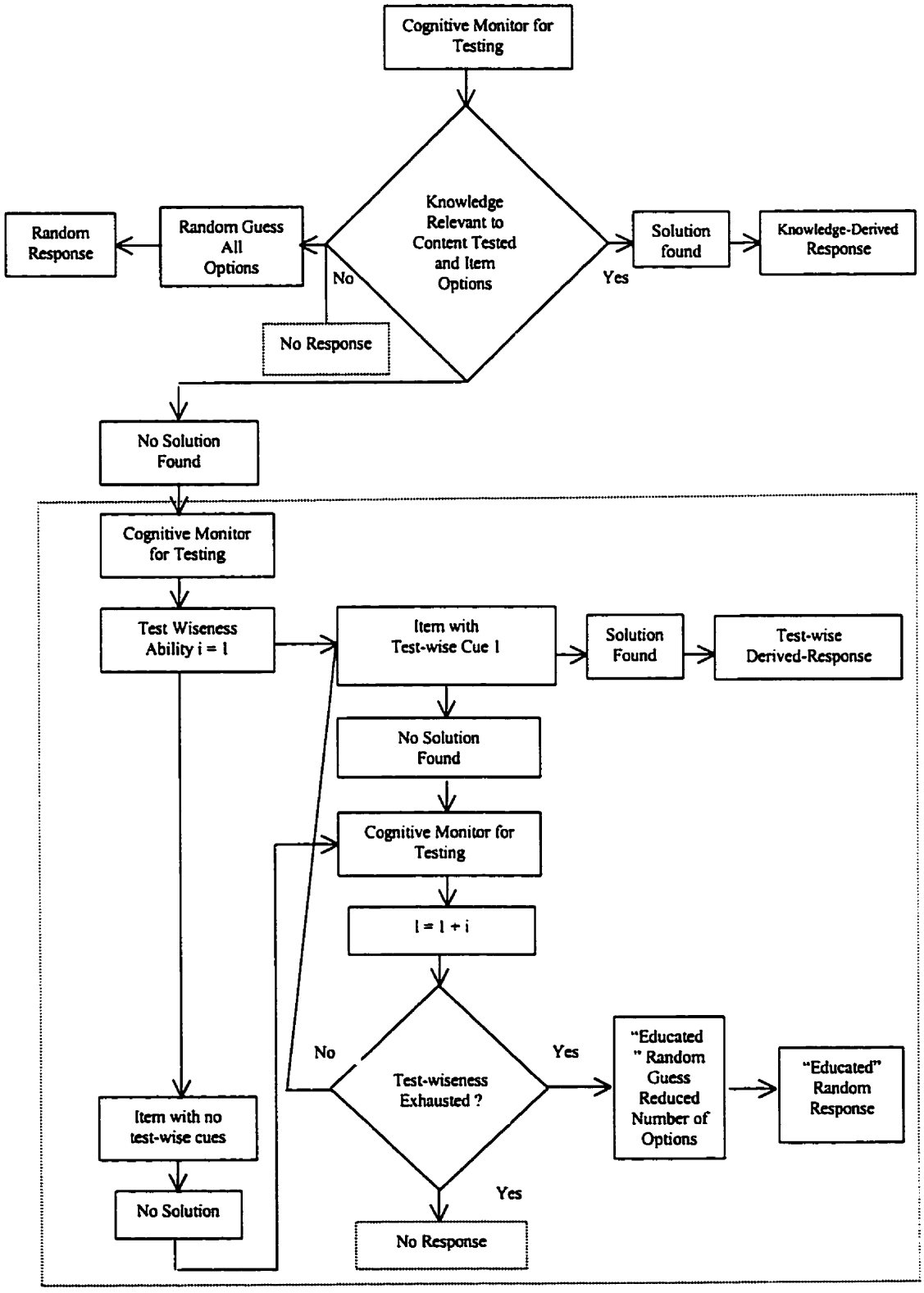

Figure 1. A model of test-taking behavior of skilled test-takers (Rogers & Bateson, 1991a).

to select on a multiple choice item.

In an attempt to validate their model, Rogers and Bateson (1991a) randomly chose and then interviewed 77 twelfth graders out of a sample of 936 students from 10 schools in British Columbia in Canada who had written the provincial school-leaving examinations as well as the Test of Test-wiseness (TTW; Rogers & Bateson, 1991a). The empirical evidence attained from a protocol analysis of the "think-aloud" data provided by these students revealed that the model presented by Rogers and Bateson (1991a) was basically consistent with the actual processes of those students when responding to 14 items specially constructed with particular item flaws (Rogers & Bateson, 1991a). Nonetheless, it is unknown yet whether this model is generalizable across other populations and other examinations containing multiple-choice items. Thus, further inquiry into examination of the generalizability of the model is in order.

The Purpose of the Research

The purpose of the present study was to examine the effects of test-wiseness on items of the Test of English as a Foreign Language (TOEFL). It involved in-depth investigations on (a) the strategies non-native speakers of English (NSE), specifically, a group of Chinese students used when responding to each TOEFL item in the multiple-choice format; (b) whether the strategies used were consistent with Rogers and Bateson's model (1991a); and (c) the extent to which test-wiseness influenced test scores. More specifically, the following questions were addressed in the study:

1. Are there any test-wise susceptible items on the TOEFL? If yes, to what extent does the presence of such items influence the test score and the interpretation of

performance?

2. What general cognitive processes do Chinese candidates employ when applying their test-wiseness, coupled with relevant partial knowledge, to respond to a test-wise susceptible item on the TOEFL?

3. Is Rogers and Bateson's (1991a) model of test-wise test taking behaviour (see p. 3 for reference) for high school seniors generalizable to Chinese population? If not, what is distinctive about the group Chinese candidates who wrote the TOEFL in terms of their test-taking behaviour?

Definition of Terms

For the purposes of this study, the definitions of following terms will be used:

Test-wiseness: It is also called test-taking strategies/skills, test-sophistication, test-familiarity, test-taking orientation, and test-wisdom. Test-wiseness is defined as "a subject's capacity to utilize the characteristics and formats of the test and/or test-taking situation to receive a high score" (Millman, Bishop, & Ebel, 1965, p.707). If an examinee possesses relevant partial knowledge as well as test-wiseness and if the test contains susceptible items, then the combination of these factors could inflate the test score. Test-wiseness, if irrelevant to the target construct to be measured, is considered as a source of error variance in test results, forming a threat to the validity of test score interpretation (Thorndike, 1951; Millman et al., 1965; Sarnacki, 1979; Prell & Prell, 1986; Rogers & Bateson 1991a).

In this study, attention was given to the four test-wiseness cues that are most frequently identified in standardized tests involved in previous studies (Metfessel & Sax, 1958; Bangert-Drowns, 1983; Benson, 1988; Slack & Porter, 1980; Smith, 1982; Rogers & Bateson, 1991a; Rogers & Wilson, 1993). Listed in the taxonomy outlined by Millman et al. (1965), these four test-wiseness cues are absurd options (ID1), similar options (ID2), opposite options (ID3), and stem option similarity (IIB4).

Test-wise: individuals who possess a substantial amount of test-wiseness and, more specifically, whose score is approximately one standard deviation above the mean on the Test of Test-wiseness (TTW; Rogers & Bateson 1991a).

Test-naive: individuals who are deficient in test-wiseness and, more specifically, who receive a score approximately one standard deviation below the mean on the TTW (Rogers & Bateson, 1991a).

Partial knowledge: knowledge possessed by an examinee which is relevant to the subject area being measured but insufficient for her/him to determine the correct answer alone when responding to a relevant test item.

Content-free items: items whose content was deliberately set up to be either nonsensical, trivial, or beyond the examinee's cognitive ability so that the correct answers could only be arrived at through application of specific test-wiseness skills, rather than through knowledge of specific subject material (Sarnacki, 1979; Bajtelsmit, 1975; Crehan et al., 1974; Slakter et al., 1970; Woodely, 1973).

Educated guessing: to eliminate one or more options as incorrect through the application of both test-wiseness skills and the partial knowledge of the test

content and then to guess randomly from among the remaining options.

The Significance of the Research

The significance of the proposed research was three-fold. First, reinforced by quantitative data and analyses, a detailed protocol analysis (Ericsson & Simon, 1993) of examinee's test performance and behaviour shed some light on the impact of test-wiseness upon performance of the TOEFL. This provided useful information for test constructors to minimize the susceptibility of items to the application of test-wiseness in the TOEFL, and for users to interpret TOEFL scores more meaningfully.

Second, since TOEFL scores are often used as an important criterion on the tertiary admission and professional certification in North America as well as many other countries where English is the language of instruction, the TOEFL has produced a substantial side effect, known as backwash effect (Hughes, 1989). Classroom pedagogy, curriculum development, language assessment, and educational policy in the ESL/ EFL community throughout the world have been shaped to "fit" the format and content of the TOEFL. Hence, the possible improvement of the TOEFL through the elimination of the items susceptible to test-wiseness may have a positive backwash on second/foreign language learning, pedagogy, and testing world-wide.

Finally, since this research was intended to further validate Rogers and Bateson's (1991a, b) model of test-wise test taking behaviour with a sample of the Chinese population, it contributed to a better understanding of test-taking behaviour and effects of test-wiseness on multiple choice items.

Scope and Delimitations of the Research

The study was focused mainly on an empirical investigation of the test-wise susceptibility of the TOEFL and on an analysis and generalization of common cognitive procedures that people usually take when responding to a TOEFL item. Other related issues such as coaching effects, practice effects, backwash effects, academic predictability of the TOEFL, second/foreign language acquisition, teachability of test-wiseness, and the dimensionality of test-wiseness were beyond the scope of the present study. Moreover, since the study was restricted to Chinese TOEFL candidates, the findings may not be generalizable to the TOEFL candidates in other countries and to other examinations.

Overview of the Dissertation

The remainder of this study is organized in the following way: Chapter II is a review of the literature, including (1) a brief description of the historical development and the current status of the TOEFL, and (2) a thorough discussion on test-wiseness and related research. The research design and data collection procedures in the proposed study are described and justified in Chapter III.

Chapter IV is divided into two sections: (1) the presentation of the overall performance of the full sample (n = 390) on the TTW; and (2) the presentation and discussions of the strategies used by the test-wise and test-naïve subsamples when responding to the TTW Interview Form.

The order of presentation of Chapter V is the same as that used in Chapter IV. The chapter is also divided into two sections: (1) the presentation of the overall performance of the full sample on the TOEFL Practice Test B (ETS, 1995d); and (2) the presentation of discussions of the strategies applied by the test-wise and test-naïve subsamples when responding to TOEFL Interview Form. The dissertation closes with a summary of the findings, implications, and limitations of the present study as well as implications for practice and for further research.

CHAPTER II

LITERATURE REVIEW

The review of the literature in this chapter is organized in two sections. In the first section, a brief description of the TOEFL program is presented. Included in this description is the historical context of the TOEFL, its format, and research related to performance on the TOEFL. Since the study was focused on effects of test-wiseness on the TOEFL, more detail is given in the second section to a review of the definition and understanding of the nature of test-wiseness. Research on correlates of test-wiseness and on the susceptibility of standardized tests to the application of test-wiseness is also summarized and reviewed.

The Test of English as a Foreign Language

The TOEFL Program

The TOEFL program is jointly sponsored by ETS, the College Entrance Examination Board (CEEB), and the Graduate Record Examination Board (GREB). The testing program is designed to evaluate the English proficiency of people whose native language is not English. It is currently required for the purpose of admission by more than 2,400 colleges and universities in the United States and Canada (ETS, 2000). It is also required as a base by various agencies and boards concerned with the accreditation and licensure of professionals from non-English speaking countries. New forms of the test are administered on a monthly basis at more than 1,275 test centres in 180 countries and areas worldwide. Approximately, 808,000 people were tested in the TOEFL program

in 1993-94 (ETS, 1995b).

The TOEFL Test

There are two formats of the TOEFL which differ in length and content. As shown in Table 1, the standard form or short form, consists of 140 multiple-choice questions. It is most likely to be administered on the dates of the "TOEFL Disclosed Test Administrations" in May, July, August, October, and December (ETS, 1995b). The term "Disclosed Test Administrations" is used when the test book is made available upon the request to any examinee who completes the test.

The long form contains from 200 to 220 items. It is made up of 140 basic items in a standard form and 60 to 80 "experimental items". It is usually 30-40 minutes longer than the standard form. The experimental items are administered mainly for the purpose of pre-calibration of the item pool of the TOEFL. Hence, they are not used to derive the examinees' scores. Accordingly, the long form, which is administered whenever the short form is not, is not released to any examinees afterwards for security reasons.

In addition, the TOEFL Test of Written English (TWE) is a short essay test designed to assess the writing ability of the candidates. The TWE is included in five TOEFL/TWE test administrations each year, usually when the short form of the test is administered (February, May, August, October, and December). The score of the TWE, however, is not added to the total TOEFL scores. Instead, it is reported separately from the scores of the three-part TOEFL in the TOEFL score report (ETS, 1995b).

Table 1.

Summary of Length and Time Allocation of the TOEFL

SECTIONS	PART TEST NAME	PART TEST TYPE	SHORT FORM		LONG FORM	
			Number	Time (Minute)	Number	Time (Minute)
Section I: Listening Comprehension			50	35*	80*	50*
Measures ability to understand spoken American English	Part A	Short conversations	30			
	Part B	Long conversations	7~8*			
	Part C	Monologues/talks	12~13*			
Section II: Structure and Written Expression			40	25	60	35
Measures ability to recognize correct standard written English	Structure	Sentence Completion	15			
	Written Expression	Error Detection	25			
Section III: Reading Comprehension			50	55	80*	70*
Measures ability to comprehend standard written English	Vocabulary	Synonym matching in reading context	50 total			
	Reading Comprehension	Reading comprehension				
Test of Written English (TWE)			1	30		
Assess the ability to compose a short essay in standard English						

Note: * The number of items or time may fluctuate from one administration to another.

Standard TOEFL Form

Introduced with the July 1995 administration, the standard form of the TOEFL, which was employed in the study, consists of three sections:

(1) Section I: Listening Comprehension comprises 50 items designed to measure the ability to understand English as spoken in North America. This section contains three parts, each of which is presented via audio-tape recorder. Part A contains 30 items. For each item, the examinee first hears a short conversation between two people. The conversation is immediately followed by a question asked by a third voice. Then, the examinee must choose the sentence that best answers the question from among four printed sentences. For each item in Part B, the examinee first hears a longer conversation after which a number of questions pertinent to the conversation are orally presented by a third person. The examinee must then choose the best answer to each question from among four printed phrases. For each item in Part C, the third part of Section I, the examinee hears a monologue, such as a simulated news broadcast, short lecture, oral presentation, or public announcement presented in segments. Immediately after each segment, several questions are asked orally by another voice. For each question, the examinee must choose the best answer from among four printed sentences or phrases. It should be mentioned that although the number of items/questions within Part B and Part C varies from one form to another, the total number of items (20) between the two parts always remains the same.

(2) Section II: Structure and Written Expression consists of 40 items developed to assess the ability to recognize language appropriate for standard written English. There are two parts in this section. Part A, Structure, includes 15 sentence-completion items. The examinee first reads each incomplete sentence and then chooses the one that best completes each sentence from among four given words or phrases. Part B, Written Expression, contains 25 error identification items. Each item is comprised of a sentence in which four words or phrases are underlined and marked A, B, C, and D. The examinee reads the sentence and then detects which one of the underlined words or phrases would not be accepted in standard written English.

(3) Section III: Reading Comprehension includes 50 items developed to evaluate the ability to read and understand a variety of general reading materials similar in topic and style to the reading materials students in North American colleges and universities are likely to encounter. The examinee reads approximately 5 short passages, each of which is followed by a number of items regarding the semantic meaning of the passage as well as specific syntactic or lexical meaning within the passage. For each item, the examinee must choose the best answer from among the four printed words, phrases, or sentences provided.

The History and Development of the TOEFL

Administration

At a 1961 conference on English language testing sponsored by the National Association for Foreign Students Affairs, the Institute of International Education, and the Centre for Applied Linguistics, the need was expressed for a systematic measure of the

English proficiency of foreign applicants to U.S. colleges and universities. Based upon the suggestions made at this conference, the National Council on the Testing of English as a Foreign Language, representing approximately 30 organizations, was established (Spolsky, 1990). With support from the Ford and Danforth Foundations in 1963, the Council commenced the groundwork toward development of a test, named the Test of English as a Foreign Language (TOEFL). This test was first administered in February 1964 and again in November 1964 and January 1965. Altogether approximately 1,800 examinees, including speakers of 71 different languages in 59 countries, took the test (Jameson & Melcolm, 1973; Oller & Spolsky, 1979; Palmer, 1965). More than 100 universities and institutions employed the results of these administrations.

In an attempt to meet the increasing needs for test development and use, CEEB and ETS assumed joint responsibility for the TOEFL program in July 1965. CEEB was responsible for promoting the use of the program by colleges and universities, while ETS was responsible for the operation of the program including construction of test forms, examinee registration, establishment of test centres, test scoring and score reporting, statistical analysis, and research. In 1966, CEEB and ETS appointed a six-member Committee of Examiners whose members included specialists in linguistics, psycholinguistics, and the teaching English as a foreign/second language. The Committee of Examiners reviewed new test forms, provided suggestions and recommendations pertinent to research, and monitored new developments in the field of second language learning and testing. In the meantime, the original National Council of the TOEFL evolved into the National Advisory Council on the TOEFL. It provided CEEB and ETS with advice regarding general policies pertinent to the development and

administration of the TOEFL as well as the interpretation and use of the scores yielded by the TOEFL.

In light of the increase in the number of graduate applicants taking the TOEFL, a new arrangement was made in 1973. Both CEEB and GREB assumed responsibility for the direction of the program, whereas ETS continued to manage the program operation. The National Advisory Council on the TOEFL was replaced by the TOEFL Policy Council as the policy-making body for the test. Rather than a general group representing several different constituencies as the earlier council was, the new Policy Council had a much more defined structure. The Committee of Examiners then became a standing committee of the Policy Council. In 1976, in an effort to cope with the growing need for research on the TOEFL, a second standing committee, the TOEFL Research Committee, was established to review research proposals and monitor research projects related to the development and use of the TOEFL. Later in 1979, the Services Committee was created to operate the administration of the TOEFL (Peirce, 1992).

The TOEFL program's growth can be seen in the marked increase, worldwide, in the number of examinees tested annually. For example, in the year 1988-89, 566,000 candidates registered to take the TOEFL; this figure rose to 675,000 in 1989-90, 741,000 in 1990-91, and 808,000 in 1994-95 (ETS, 1990, 1991a, 1992, 1995b).

Test Format

The number of subtests included in the TOEFL has been revised twice since 1963. From 1963 to 1976, the TOEFL consisted of five parts (200 multiple-choice items): (1) Listening Comprehension (50 items, 40 minutes), (2) English Structure (40 items, 20

minutes), (3) Vocabulary (40 items, 15 minutes), (4) Reading Comprehension (30 items, 40 minutes), and (5) Writing Ability (40 items, 25 minutes). In 1976, Pike found from his study (subsequently published in 1979) that Section (2) English Structure was highly correlated with Section (5) Writing Ability, and that Section (4) Reading Comprehension was highly correlated with Section (3) Vocabulary. Based upon Pike's findings, the five-part TOEFL test was re-organized into a three-part test (150 items) in 1976. The original English Structure and Writing Ability subtests were consolidated into one section, Structure and Written Expression (40 items), and the original Reading Comprehension and Vocabulary subtests were merged into another section, Vocabulary and Reading Comprehension (60 items). These changes in format resulted in a reduction in the number of items (from 200 items to 150 items) and in the testing time (approximately from 2 hours and 20 minutes to 1 hour and 45 minutes). Since the equating system and score scale remained the same, total scores on the three-part test could be interpreted in the same way as the total scores on the old five-part test. However, the subscores for each section were not comparable.

Since the use of the three-part test, much criticism was focused on the decontextualized nature of Part A (10 short statement items) in Listening Comprehension and Part A (30 vocabulary items) in the Vocabulary and Reading Comprehension subtests. As a result, the three-part test was revised again into the current standard form of the TOEFL introduced in July, 1995 (see pp. 10-12).

In July 1998, the computer-based TOEFL test was introduced in many parts of the world. The test consists of four sections, i.e., Listening (30-50 items), Structure (20-25 items), Reading (44-55 items), and Writing (one topic in 30 minutes). The Listening and

Structure sections are computer-adaptive, where the items are chosen by the computer based on how an examinee has answered the previous items. The writing section is mandatory and the examinees are required to compose an essay on a topic selected by the computer from a pool of topics. The rating of the essay approximately constitutes one half of the Structure/Writing scaled score. Scores on the computer-based test are reported on the new scales, which contains Listening (0 to 30), Structure/Writing (0 to 30), Reading (0 to 30), and a total score (0 to 300). Given a drastic change in the content and format of the test, these new scale scores, as ETS (2000) suggested, are not directly comparable with the scale scores on the paper-based test, which is still administered in many countries including China. In 1999, Jiang (1999) conducted research using the data collected from 8,387 examinees between November 1997 and March 1998 to determine the comparability of the scores earned on the computer-based TOEFL test with those earned on the paper-based test. As a result of the study, concordance tables listing equivalent scores and score ranges for both the computer-based test and paper-based test were developed. Using these concordance tables, scores obtained from the two types of the tests may be compared.

Currently, both the paper-based and the computer-based TOEFL tests are administered throughout the world. The former is gradually being phased out as the latter is phased in. More than 300,000 people registered to take the computer-based TOEFL test beginning in July 1998 through 1999(ETS, 2000).

Guidelines for the Construction of the TOEFL

The following guidelines have been used to direct the development of the various forms of the TOEFL since 1963: (a) contrastive analysis should not serve as a basis for test construction, (b) testing of reading comprehension should not involve difficult vocabulary, (c) language should be presented in a realistic context, (d) the test should be standardized in relation to a large population of nonnative English speakers, and (e) both discrete and integrative skills should be tested (Angelis, 1979; Oller & Spolsky, 1979).

However, over the past 30 years, these guidelines have been challenged and criticized by linguists and language teachers. For example, many have argued that the test should be standardized with reference to native speakers rather than with reference to nonnative speakers, especially if the test is designed to determine an applicant's readiness to study in North America (e.g., Oller et al., 1979). In practice, numerous studies have demonstrated that not all sections in the TOEFL (e.g., Structure and Written Expression) have always focused on tasks that native speakers do well (Angoff & Sharon, 1971; Angelis, 1977; Clark, 1977; Johnson, 1977). In 1971, Angoff and Sharon found that nonnative speakers outperformed their native speaker counterparts on 21% of the items in the Writing Ability section which tapped examinee's knowledge of grammatical forms. Angelis (1977) and Clark (1977) found that approximately 20% of items in the Structure and Written Expression section and in the Reading Comprehension section were more difficult for native speakers than expected. The occurrence of the unexpected difficult items may be attributed partly to the native speakers' lack of grammatical skills, partly to the abstractness and lexical rarity in the vocabulary section of the TOEFL, and partly to the need to make complicated judgements and inferences in the Reading Comprehension

subtest of the TOEFL (Angoff & Sharon, 1971; Angelis, 1977; Clark, 1977).

Use and Interpretation of TOEFL Scores

The TOEFL is designed to measure the English proficiency of people whose native language is not English with the ultimate purpose of providing useful information for making admission decisions. Regarding the use of the TOEFL for admission purposes, a prevalent debate over three decades has concentrated on the prediction of a foreign student's academic success on the basis of the TOEFL score. Numerous studies have been completed to examine the TOEFL's relation to student's grade point average (GPA) and other common measures of aptitude, intelligence, and achievement (Hale et al., 1984; Graham, 1987). However, partly because of the complex relationships between academic success and its predictive variables, and partly because of small-size samples or samples with restricted range (e.g., samples including only examinees above the cut-off score), the conclusions are inconsistent and contradictory. Some researchers have argued that it is not appropriate to employ TOEFL scores to predict academic success in terms of GPA (e.g., Hwang & Dizney, 1970; Sharon, 1972; Wilcox, 1975; Gue & Holdaway, 1973; Light et al., 1987; Ayers & Quattlebaum, 1992). Others have maintained that the TOEFL is one of the major predictors (if not the sole predictor) of academic performance (e.g., Burgess & Greis, 1970; Heil & Aleamoni, 1974; Ayers & Peters, 1977; Odunze, 1982; Ho & Spinks, 1985). Hence, further research on this issue with large and unrestricted samples is clearly in order.

In light of this research, ETS itself has gradually shifted its standpoint from supporting to opposing the use of the TOEFL alone as a predictor of academic performance. For example, the 1968 TOEFL manual (ETS, 1968) referred to several studies in which a positive relation between TOEFL scores and GPA was noted. It was asserted that the findings supported the use of the TOEFL for prediction purposes. Nonetheless, such an assertion was no longer found in the 1973 TOEFL manual (ETS, 1973). Instead, a low-profiled view was presented: although the results of positive correlations between TOEFL scores and GPA could be expected if the TOEFL scores were used appropriately, such positive correlations were usually so low as to be of little practical use in the admission process. Following their review of more than 100 studies that examined the use of the TOEFL in admission processes, Hale, Stansfield, and Duran (1984) questioned the validity of the TOEFL as a predictor of success in graduate school as measured by GPA. As a result, ETS (1983) completely abandoned its initial viewpoint and strongly cautioned institutions against the use of the TOEFL scores either for predicting academic performance or as the sole basis for admissions decisions. In 1997, ETS reiterated its standpoint:

> It should be pointed out that the TOEFL test is only a measure of general English proficiency. It is not a test of academic aptitude or of subject matter competence, nor is it a direct test of English speaking or writing ability. (ETS, 1997, p. 25)

Another issue often raised is that, due to error of measurement, TOEFL scores should not be regarded as a perfect measure. Rigid cut-off scores, as suggested in the TOEFL manual (1995c), should not be used for admissions decisions without taking into account error of measurement. In practice, however, the issue is largely ignored. For

example, Yalden (1978), in her survey, found the consistent use of a fixed score on the TOEFL as a condition for acceptance by almost all Canadian universities, without reference to error of measurement or the use of other additional predictors.

Research on the TOEFL

Given its prominence in the area of second/foreign language testing and acquisition, the TOEFL has been the subject of continual research ever since it was developed in 1963. In addition to the research on the predictive efficiency of the TOEFL briefly reviewed above, many other issues have been addressed in the following areas: validity and reliability of the inferences drawn from the TOEFL score, the construction and use of the TOEFL, examinee performance, and psychometric and statistical technology involved in the TOEFL construction and item analysis (ETS, 1995a; Hale, et al., 1984). However, for the purposes of the study, only a few related aspects are addressed below.

Construct Validity of Inferences Based on the TOEFL Scores

Construct validity of the evidence for inferences made using TOEFL scores has been collected through empirical research on the TOEFL 's relation to other similar standardized tests of English proficiency, such as the Michigan Test of English Language Proficiency (MTELP) and the International English Language Test System (IELTS) (Hale et al, 1984; Dizney, 1965; Abadzi, 1976; Pack, 1972; Buell, 1992; Geranpayeh, 1994). Also relevant to construct validity are investigations of the relationships between the TOEFL and many other direct or indirect measures of English language proficiency, including the cloze test, dictation, oral interviews, essay writing, and instructor's rating

(Hosley & Meredith, 1979; Scholz & Scholz, 1981; Clark & Swinton, 1980; Pike, 1979; Warner, 1982; Irvine et al., 1974). Based on the consistent findings of these studies that the TOEFL was moderately to highly correlated with these measures of English proficiency, it seems that valid inferences can be made using the TOEFL scores for assessing non-native speaker's English language proficiency (Hale et al., 1984; ETS, 1993b).

Nonetheless, contrary arguments have been made, especially by ESL teachers and linguists. They have argued that the three-part TOEFL is an incomplete language test because it was designed to measure only the receptive aspect (listening and reading) of English language proficiency. Consequently, the TOEFL fails to assess productive language abilities such as speaking and writing, both of which are essential for success in academic studies in colleges and universities (Woodford, 1978; Choy & Davenport, 1986; Traynor, 1985; Raimes,1990). Furthermore, this incompleteness may bring about a negative backwash effect on ESL instruction. Since speaking and writing abilities are not tested, they are often neglected in teaching. Instead, pedagogical emphasis is given to rote learning of English grammar and vocabulary (Choy, 1986; Traynor, 1985; Raimes, 1990).

Relationship between the TOEFL and Examinee's Characteristics

A substantial number of research studies have been conducted to examine whether there are any significant and meaningful differences in test performance on the TOEFL due to language or cultural background, gender, age, educational level (graduate vs. undergraduate), and disciplines (ETS, 1995a). Across these studies, the findings are equivocal and the conclusions contradictory. On one hand, independent researchers such

as Farhady (1982) demonstrated with some empirical evidence that a significant discrepancy in test performance on the TOEFL was closely related to examinee characteristics including gender, educational level, nationality, and language/cultural background. On the other hand, most of the researchers working for ETS argued that, although significant differences in test performance on the TOEFL may exist between or among languages groups, sexes, age, and educational levels, such differences usually can be seen to be of relatively small magnitude when expressed on the TOEFL scale and, consequently, they are not very meaningful in practice (Hale et al., 1984; Hale, 1988; Alderman & Holland, 1981; Wilson, 1982; Swinton & Powers, 1980; Oltman et al., 1988; Brown, 1993). Further research is needed to clarify which examinees' characteristics are meaningfully related to performance on the TOEFL and what the consequences are for the interpretation and use of the TOEFL.

Effects of Test-wiseness on the TOEFL

Although Oltman, Stricker, and Barrows (1988) supported ETS's claim that the TOEFL is relatively insensitive to examinees' characteristics, they noted a prevalent phenomenon in their research findings that low-scoring examinees from certain language groups tended to limit their success on the test by declining to respond to every item. Yet examinees are encouraged to answer every item and told that the guessing is not penalized in the test instructions. One of the possible explanations for this finding, suggested by Oltman et al. (1988), is that these examinees might lack relevant test-wiseness skills.

Other than a few studies (e.g., Mullen, 1978; Des Brisay & Ready, 1991; Wilson, 1987; Palmer, 1984) in which the effects of instruction and practice on the TOEFL were addressed as a side question, there appears to have been no investigation of the effects of test-taking skills or test-wiseness on the test. Although the TOEFL test is well constructed and rigorously reviewed (Peirce, 1992), it would be scientifically naive to believe that this test is immune to test-wiseness inasmuch as substantial research has consistently indicated that test-wiseness is a pervasive factor affecting all kinds of tests on any subject matter, regardless of whether they are teacher-made or standardized (Sarnacki, 1979; Mehrens & Lehmann, 1973; Oakland, 1972; Gaines & Jongsma, 1974; Gross 1976; Callenbach, 1973; Rogers & Bateson, 1991a, b; Rogers & Wilson, 1993; Rogers & Yang, 1996). In this regard, it is justifiable to empirically examine the relationship between test-wiseness and the TOEFL. This justification is further strengthened by the discussion in next section on test-wiseness.

Test-Wiseness

Test-wiseness, also called test-taking skills/strategies, test sophistication, test-familiarization, test-taking orientation, or test-wisdom (Anastasi, 1976; Erickson, 1972; Sarnacki, 1979), is a complex phenomenon that is reflected in test performance and accounts for some systematic variance in test scores. In attempting to demonstrate the construct validity of test-wiseness, the literature reviewed in this section is organized in the following three subsections: (1) definition of test-wiseness, (2) the construct of test-wiseness, and (3) the susceptibility of standardized tests to test-wiseness cues.

Definitions of Test-Wiseness

Thorndike (1951) is credited as being the first to recognize test-wiseness as a persistent factor that can influence test performance (p. 568). His description of test-wiseness follows:

> Shrewdness with regard to when to guess, and a keen eye for secondary and extraneous cues are likely to be useful in a wide range of tests, particularly those that are not well constructed. (p. 569).

After describing characteristics of test-wiseness, Thorndike went on to suggest that "it usually represents systematic invalid variance serving systematically to reduce the validity of the test" (p. 569). Thus, he set the stage for the controversy that exists today: *is test-wiseness a relevant or irrelevant influence upon test performance?*

Despite Thorndike's suggestion that the validity of the interpretation of a test score may be compromised by the influence of test-wiseness, it was not until 1964 that the first empirical study of test-wiseness was conducted. Based on Thorndike's brief description of test-wiseness, Gibb (1964) developed and validated a scale to measure test-wiseness. In his study, Gibb referred to test-wiseness as "the ability to react profitably to the presence of secondary cues in a test" (p. 5). Like Thorndike, Gibb felt that, given individual differences in test-wiseness, it was a source of systematic error variance that could jeopardize the valid interpretation of a test score.

In the next year, Millman, Bishop, and Ebel (1965) published what has become the most frequently quoted definition of test-wiseness: "a subject's capacity to utilize characteristics and formats of the test and/or test taking situation to receive a high score. Test-wiseness is logically independent of the examinee's knowledge of the subject matter for which the items supposedly measures" (p. 707). This definition, as interpreted by

Sarnacki (1979), suggested that test-wiseness encompasses both genetic deficiencies inherent within test construction and item format, and a cognitive ability (or abilities) that an examinee may employ to improve a test score in any testing situation regardless of the content area being measured.

Immediately following Millman et al.'s (1965) seminal work, several studies on the influence of test-wiseness were conducted, with only minor changes in the definition of test-wiseness. For example, Oakland (1972) defined test-wiseness as "the ability to manifest test-taking skills which utilize the characteristics and format of a test and/or test-taking situation in order to receive a score commensurate with the abilities being measured" (p. 355). Likewise, Smith (1980, 1982) viewed test-wiseness as the interaction between characteristics of a test-taker and characteristics of test format. Diamond and Evans (1972) referred to test-wiseness as "the ability to respond advantageously to multiple choice items containing extraneous clues and to obtain credit on these items without knowledge of the subject matter" (p. 135). Similarly, Williams and Dolly (1983) defined test-wiseness as the "ability of the test-taker to perform at better than chance level on a multiple choice test no matter what the content being tested" (p. 2). Sarnacki (1979), however, questioned the apparent restriction of test-wiseness to multiple-choice items, asserting that the influence of test-wiseness is present in other types of test items as well.

More recently, Rogers and Bateson (1991a, b) questioned the simple interpretation that test-wiseness and subject matter knowledge are independent. Based upon the findings of their research, they suggested:

> Thus it appears that effect application of test-wiseness reasoning strategies is dependent on some partial knowledge. This partial knowledge, although inadequate to respond to the test item solely on the basis of this knowledge, is sufficient when coupled with knowledge of the test-wiseness principles to increase the probability of correct responding to items susceptible to test-wiseness. Students with low content knowledge but test-wise knowledge and students with partial knowledge but low test-wise knowledge will perform less well than students who possess both on such items (p. 210).

Rogers and Bateson's (1991a) suggestion opened another new controversy on the construct of test-wiseness: *is test-wiseness dependent or independent upon the subject matter knowledge being measured?*

Construct of Test-wiseness

Taxonomy of Test-wiseness Components

In order to further explicate the construct of test-wiseness, Millman et al. (1965) developed a Taxonomy of Test-wiseness Principles (pp. 711- 712) which has served as a basic conceptual framework for the construct of test-wiseness and studies conducted subsequent to the work of Millman et al. As shown in Table 2, this Taxonomy is organized into two major parts. Part I contains elements independent of the test maker or test purpose and applicable in most testing situations. The first two subcategories in Part I consist of strategies which, if applied, may help examinees avoid losing marks for reasons other than lack of the content knowledge being measured. The last two subcategories in Part I are composed of elements or strategies which allow examinees to gain extra credits beyond what they would otherwise have received on the basis of sure and full knowledge of the specific content area being tested.

Table 2.

Taxonomy of Test-wiseness Principles

I. Elements independent of test constructor or test purpose.

A. Time-using strategy.

1. Begin to work as rapidly as possible with reasonable assurance of accuracy.
2. Set up a schedule for progress through the test.
3. Omit or guess at times (See I. C. and II. B.) which resist a quick response.
4. Mark omitted items, or items which could use further consideration, to assure easy relocation.
5. Use time remaining after completion of the test to reconsider answers.

B. Error-avoidance strategy.

1. Pay careful attention to directions, determining clearly the nature of the task and the intended basis for response.
2. Pay careful attention to the items, determining clearly the nature of the question.
3. Ask examiner for clarification when necessary, if it is permitted.
4. Check all answers.

C. Guessing strategy.

1. Always guess if right answers only are scored.
2. Always guess if the correction for guessing is less severe than a "correction for guessing" formula that gives an expected score or zero for random responding.
3. Always guess even if the usual correction or a more severe penalty for guessing is employed, whenever elimination of options provides sufficient chance of profiting.

D. Deductive reasoning strategy.

1. Eliminate options which are known to be incorrect and choose from among the remaining options.
2. Choose neither or both of two options which imply the correctness of each other.
3. Choose neither or one (but not both) of two statements, one of which, if correct, would imply the incorrectness of the other.
4. Restrict choice to those options which encompass all of two or more given statements known to be correct.
5. Utilize relevant content information in other test items and options.

Table 2. (Continued)

II. Elements dependent upon the test constructor or purpose.

A. Intent consideration strategy.

1. Interpret and answer questions in view of previous idiosyncratic emphases of the test constructor or in view of the test purpose.
2. Answer items as the test constructor intended.
3. Adopt the level of sophistication that is expected.
4. Consider the relevance of specific detail.

B. Cue-using strategy.

1. Recognize and make use of any consistent idiosyncrasies of the test constructor which distinguish the correct answer from incorrect options.

    a. He makes it longer (shorter) than the incorrect options.
    b. He qualifies it more carefully, or makes it represent a high degree of generalization.
    c. He includes more false (true) statements.
    d. He places it in certain physical positions among the options (such as in the middle).
    e. He places it in a certain logical position among an ordered set of options (such as the middle of the sequence).
    f. He includes (does not include) it among similar statements, or makes (does not make) it one of a pair of diametrically opposite statements.
    g. He composes (does not compose) it of familiar or stereotype phraseology.
    h. He makes it grammatically inconsistent with the stem.

2. Consider the relevancy of specific detail when answering a given item.
3. Recognize and make use of specific determiners.
4. Recognize and make use of resemblances between the options and an aspect of the stem.
5. Consider the subject matter and difficulty of neighbouring items when interpreting and answering a given item

---

(Millman et al., 1965, p. 711-713)

Part II of the Taxonomy contains elements or strategies that may be beneficial when the examinee has knowledge of particular test making behaviours or particular testing practices due to past experience with tests similar in purpose and format. As in Part I, the elements in the first subcategory help examinees avoid losing points, whereas the elements in the second subcategory facilitate examinees in gaining extra credits (see Table 2 for details).

Given its comprehensive and concise nature, Millman et al.'s (1965) Taxonomy has been adopted as a blueprint for subsequent research studies in this area. In return, many of these studies have further validated and expanded the Taxonomy. For example, two additional test-wiseness elements, both related to ID4 and ID5 and both based on the notion of convergence have been suggested in the literature. Smith (1980, 1982) demonstrated that the correct answer to some multiple-choice items could be determined by looking only at the convergence among options. This strategy takes advantage of a common practice in developing multiple-choice items: the distracters usually possess various degrees of correctness to the stem and to the correct option so as to be plausible. Smith (1982) hypothesized that a test-wise examinee would select the option that was some way related, i.e., converged, to each of the other options. Presented below is one of his examples to illustrate the notion of convergence. The options are provided first, followed by a rearrangement of the options in a two-facet chart: time by food. The correct answer is A. before breakfast, an option in the cell overlapping the two facets.

Item Options

A. before breakfast
B. on a full stomach
C. with meals
D. before going to bed

Classification of Options by Food and Time

	Time	
	A. before breakfast	D. before going to bed
Food	B. on a full stomach C. with meals	

Figure 2. Illustration of option convergence (Smith, 1982, pp. 211-220)

Smith tested his hypothesis using a group of adult examinees. These examinees were asked to select what they believed to be the correct answer from among the response options without providing them with the stem. It turned out that their answers converged on the correct option at a greater-than-chance level. Likewise, in another case, a group of high school students trained in the convergence strategy were found to significantly outscore their counterparts who received instruction on other test-taking skills on a SAT verbal subtest (Smith, 1982). Although the susceptibility to the convergence strategy, as Smith (1982) argued, does not necessarily mean a major flaw in multiple choice items, functionally, it may cue shrewd examinees in much the same way as flawed items do.

Powers and Leung (1995) demonstrated another type of convergence strategy used by high school seniors when responding to multiple choice items referenced to reading passages in the SAT. The strategy involves choosing answers on the basis of consistency among the questions and their options in the set of items for the reading

passage. Using this strategy and/or some other strategies, these high school seniors were found to be able to attain scores that exceeded a chance level when only given the questions and options but not the reading passages.

Models of Test-wise Test Taking Behaviour

In an attempt to attain a complete understanding of the construct, considerable attention has been given to the general cognitive processes test-wise examinees go through when actually applying test-wiseness to respond to an item, particularly a multiple choice item. Smith (1980, 1982) first proposed a model of test-taking behaviour which reflected various routes a skilled examinee may take to determine which option to select on a multiple choice item. As demonstrated in Figure 3, there are four components in this model:

1. a cognitive monitor that determines which skills and abilities are going to be involved to respond to an item;
2. the individual's abilities and skills related to the trait being assessed;
3. personal knowledge of relevant test-taking strategies; and
4. response and refinement of relevant test-wiseness strategies.

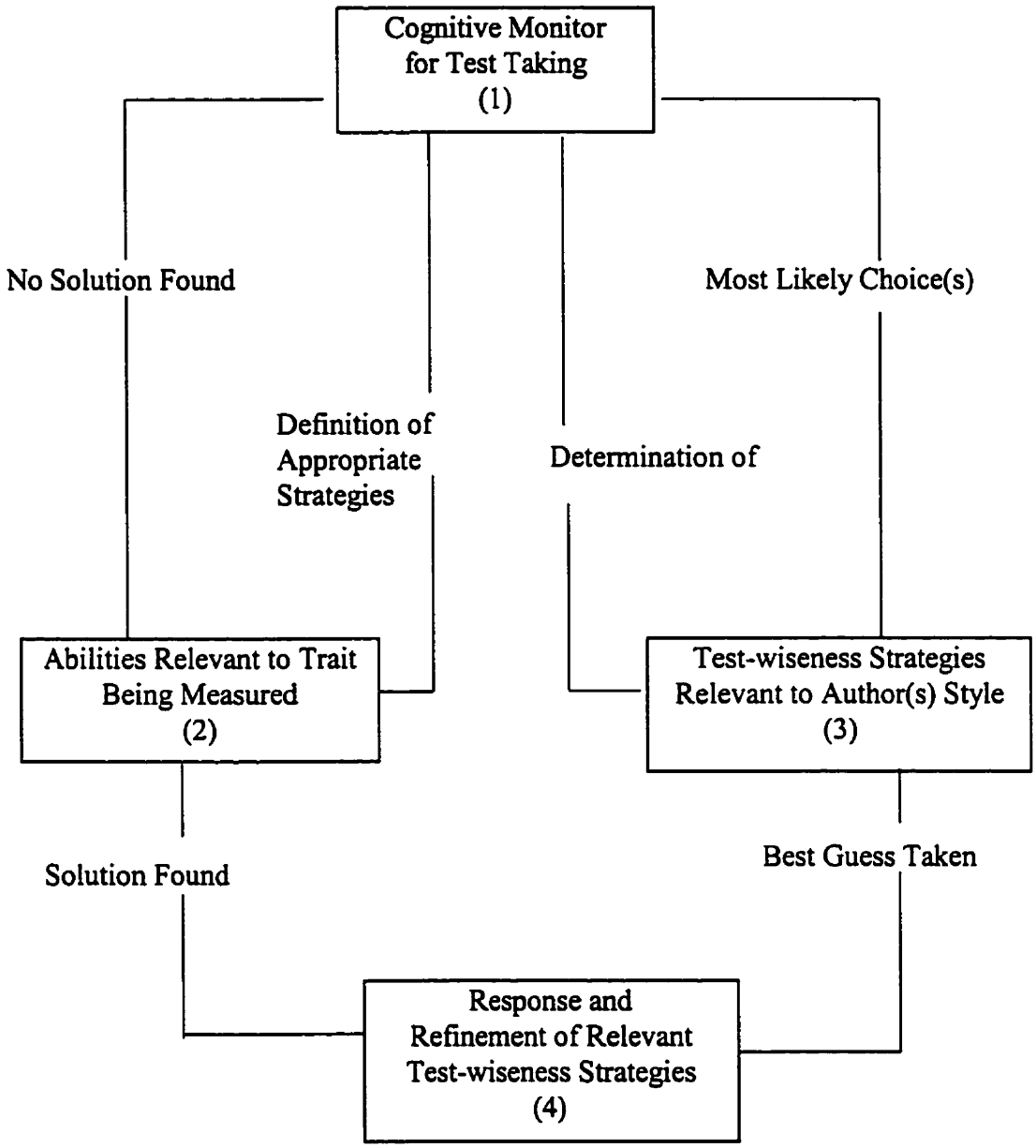

Figure 3. Smith's Model of Test Taking Behaviour of Skilled Test Takers (Smith, 1980)

Based upon Smith's model (1980) as well as the work of Brown (1980, 1987), Flavell (1979), and Schuell (1986), Rogers and Bateson (1991b) proposed and validated a revised model of test-wise test-taking behaviour. As illustrated in Figure 1 (see p. 3), Rogers and Bateson's model retains the four major cognitive components in Smith's model. In addition, it is much more detailed, complete, and sophisticated. In their model, it is explicitly demonstrated how construct-irrelevant easiness (Messick, 1989) is achieved through the combination of test-wise susceptible items, the possession of test-wiseness skills, and partial knowledge of the content tested.

In light of the defined process postulated in this model, the examinee initially reads the stem of a multiple-choice item and then attempts to identify what he or she believes to be the correct answer from among the options provided using knowledge and skills relevant to the perceived content being measured. If the answer is not found, a test-naive examinee would probably either randomly guess from among the options or simply omit the question entirely. In contrast, a test-wise person would tend to apply his or her set of test-wiseness strategies together with his or her partial knowledge pertinent to the content assessed, working cyclically for a test-wiseness element-item cue match. When such a match is found, the cycle is terminated and a test-wise response is recorded. In the case of no match, due to the absence of test-wiseness cues or to the exhaustion of all the test-wiseness strategies he or she possesses, this person will probably make an "educated" random response as a last attempt (i.e., eliminate one or more that appear to be incorrect options with their test-wiseness skills and partial knowledge of the test content and then guess randomly from among the rest). According to this model, other characteristics of the test-wise examinee in addition to knowledge of the subject matter being measured

form another source of variance in the final test score.

The empirical evidence attained from protocol analysis of 77 high school students chosen from a representative sample of 936 twelfth graders from British Columbia in Canada revealed that the model presented by Rogers and Bateson (1991a) was concordant with the actual processes of these students when responding to 14 items specially constructed with particular item flaws (Rogers & Bateson, 1991b). Therefore, Rogers and Bateson's (1991b) model was adopted as the operational framework for the present study. The research focus was placed on studying the general cognitive processes of the test-wise and test-naive examinees when responding to the TTW and test-wise susceptible items on the TOEFL.

Measures of Test-wiseness

The first comprehensive measure of test-wiseness was developed by Gibb (1964). His test consisted of 70 novel history multiple-choice items designed to measure 7 test-wiseness elements (10 items for each element). The seven elements included (1) options editorially similar to a word in the stem (stem-option cues); (2) absurd options; (3) options containing specific determiners (e.g., all, never, and always); (4) precision or qualification of answer (precise option); (5) longer correct options; (6) options containing grammatical cues; and (7) relevant content information in one item cues correct option in another option (item giveaway). Gibb reported that the internal consistency (KR-20) for the total score, computed from a sample of 193 college students, was .90 for a trained group of undergraduates (n = 101) and .72 for an untrained group (n = 92). Overall, the trained group scored significantly higher ($p < .05$) than the untrained group. Sarnacki

(1979), in his review on test-wiseness, declared Gibb's test to be the best available measure of test-wiseness. In effect, as the initial measure of test-wiseness, Gibb's (1964) test served as a guideline for the development of test-wiseness instruments in later studies.

Millman (1966) developed four tests to measure four different aspects of test-wiseness elements for use with high school and college students. The four tests are described below:

> Test A measured the ability to identify inconsistencies in options. The correct option was always (1) longer than the distracters (IIB1a), (2) qualified more carefully or contained a higher degree of generalization (IIB1b), (3) in the logical middle of an order set of options (IIB1e), or (4) one of a pair of diametrically opposite options (IIB1f) (Millman, 1966, pp. 9-13).
>
> Test B measured the ability to make use of resemblances between the options and the stem (IIB4) (pp. 13-14).
>
> Test C measured the ability to determine when it is profitable to guess (IC3) (pp. 14-16).
>
> Test D measured the ability to apply the first four of the deductive reasoning strategies (ID1, ID2, ID3, and ID4) (pp. 16-18).

All test items were "content-free" to rule out the possibility of examinees correctly responding to an item by the application of relevant knowledge alone. Each of the four tests consisted of an equal number of items from English literature, mathematics, history, political science, and biology.

The mean value of the internal consistencies (split-half correlation corrected with the Spearman Brown formula) of the four tests was .53 for the high school sample and .38 for the college sample. The mean discriminations vary from .31 to .44 across the four tests. In terms of validity, Millman (1966) suggested that the way the tests were constructed constituted "major evidence for the validity of the two tests" (p. 19). Using the multitrait-multimethod approach (Campbell & Fiske, 1959), Millman found that subject area was not influential across the four tests.

Woodley (1973) and Bajtelsmit (1975) developed test-wiseness tests for adults. Woodley constructed a test of 30 content-free items to measure the attainment of absurd options (ID1), similar options (ID2), and the stem-option cue (IIB4). Based on a sample of 259 adults students enrolled in an insurance institute, internal consistencies (Cronbach alpha) were .52 for ID1, .44 for ID2, .63 for IIB4, and .73 for the total test.

Bajtelsmit's (1975) Test of Obscure Knowledge (TOOK) contained five 10-item subtests designed to assess specific determiners (IIB3), item giveaway (ID5), and the three test-wiseness elements (i.e., ID1, ID2, and IIB4) measured by Woodley's test (1973). The values of internal consistency (Cronbach's alpha), computed from a sample of 56 undergraduates, ranged from .49 (IIB4) to .77 (IIB3).

Slakter et al. (1970) generated a test to measure four test-wiseness elements: (1) absurd options (ID1), (2) similar options (ID2), (3) specific determiners (IIB3), and (3) stem-option cues (IIB4). The test was comprised of four subtests, each of which was design to measure one of the four elements, respectively. The test was then administered to 1,070 students in grades 5-11. The median internal consistency (KR-20) for the test scores was .44 across the 7 grade levels and .63 for grades 9 through 11. The median

internal consistencies were .25 for ID1, .46 for ID2, .08 for IIB3, and .31 IIB4 (p. 120). Similar results were found in a replication of this study in which the same test was administered to another sample of 1,291 students in grades 5-11 (Slakter et al., p. 120).

Allan (1992) constructed a 33-item test for ESL students to measure similar options (ID2), grammatical cues (IIB1h), stem-option cues (IIB4), and item giveaway (ID5). The internal consistencies (KRB20) for the four subscales, determined from the responses of 51 first-year college students in Hong Kong whose first language was Cantonese, were, respectively, .19, .71, .37, and .66. The low reliabilities for the similar-option and stem-option subtests, according to Allan, may be attributed to the Hong Kong students' inadequate English language ability to recognize synonyms and analyse compound words (pp. 108-109).

Rogers and Bateson (1991a) generated a 24-item test for use with high school students to assess four test-wise strategies most frequently used by students in standardized tests. These strategies include three deductive reasoning strategies ---- ID1, ID2, and ID3 ---- and one cue using strategy ---- IIB4. The items were evenly distributed across these four strategies and approximately evenly distributed across the four major Grade 12 courses ---- English, algebra, history, and biology. The content of the items was either obscure or, in the case of some of the incorrect options of deductive items, very familiar. Computed from a sample of 936 Grade 12 students, the internal consistency (Hoyt 1941) for the total test score, was .37 and for four subscales (i.e., ID1, ID2, ID3 and IIB4) were .22, .36, .16, and .24, respectively. To further assess validity, the 36 highest scoring students and 41 lowest scoring students were interviewed and asked to "think aloud" when responding to a set of sample items selected from the Test of Test-wiseness

(TTW) they wrote before. The student responses revealed that high scoring students applied the test-wiseness strategies to a much greater extent than their low scoring counterparts. It was further revealed that, when responding to a test-wise susceptible item, most of students who lacked sufficient content knowledge to determine the correct option with certainty tended first to eliminate one or more options as incorrect with both their partial knowledge and test-wiseness skills and then to randomly guess from among the remaining options. This typical behaviour of "educated guessing" helps to explain the low values of internal consistencies across various measures of test-wiseness mentioned above.

In sum, a number of tests have been constructed since Gibb's (1964) first endeavour to measure the construct of test-wiseness. Developed in a similar way, these tests vary with respect to the test-wiseness elements assessed and the age groups to whom the tests were administered. Of all elements included in various measures of test-wiseness, the most frequently tested and also most frequently used by examinees were absurd options (ID1), similar options (ID2), opposite options (ID3), specific determiners (IIB3) and stem-option cues (IIB4) (Millman, 1966; Slakter, et al., 1970; Woodley, 1973; Bajtelsmit, 1975; Allan, 1992; Rogers & Bateson, 1991a; Rogers & Wilson, 1993; Morse, 1994). Based on the results from the validation of all these measures of test-wiseness, it was found that the estimated reliabilities (internal consistencies) were usually low, particularly for subtests involving the elements (e.g., deductive reasoning strategies) which did not always lead to a specific answer. The low values may be explained partly by the small number of items included and partly by examinee's educated guessing test-taking behaviours (Rogers & Bateson, 1991a; Rogers & Yang, 1996).

The Nature of Test-wiseness

1. Generality Versus Specificity. One major dispute in the study of the nature of test-wiseness is focused on the generality or specificity of this construct. Thorndike (1951) postulated that test-wiseness was (1) a lasting general trait in relation to an individual's ability to guess strategically and identify cues, and (2) a lasting specific trait associated with particular test types or item formats. Congruent with Thorndike's perspective, Stanley (1971) considered test-wiseness as a persistent attribute. Crehan et al. (1974) found that test-wiseness was a stable characteristic over grade levels 5 through 11 regardless of the subject area tested. Basically, these scholars contended that test-wiseness is best interpreted in terms of the individual's abilities, traits, or states, rather than the characteristics of tests.

In contrast, opponents argued, with the support of empirical evidence, that since effective application of test-wiseness relies heavily on items that are susceptible to test-wiseness and on inherent limitations of certain test formats, test-wiseness should be viewed, not as a general trait of an individual examinee, but rather as a psychometric idiosyncrasy of tests specific to the cues in flawed items (Diamond & Evans, 1972).

Nevertheless, as pinpointed by Sarnacki (1979), these two viewpoints are not mutually exclusive. Instead, they are strongly interdependent and represent the different attributes of test-wiseness. Although both viewpoints provide necessary information pertaining to test-wiseness, neither view alone is sufficient in understanding and explaining the construct which actually encompasses both the method of measurement and the characteristics of the examinee. Millman et al. (1965) attempted to demonstrate this synthesis through their Taxonomy of Test-wiseness Principles. In their view, the test

taking skills they outlined are considered to be general with respect to the academic subject of the test, but specific when responding to a specific test item. Furthermore, in terms of specificity, these test-wiseness skills can be divided into two components: (1) strategies dependent upon the examinee, and (2) strategies dependent on the test constructor and/or test purpose. To date, Millman et al.'s suggestion is widely accepted (Sarnacki, 1979; Rogers & Bateson, 1991a, b; Rogers & Yang, 1996).

Nonetheless, in terms of specificity of test-wiseness, the nature and structure of the construct is not yet well understood. Logically, if test-wiseness is a general trait, then the separate test-wiseness skills should be highly correlated. If not, it is plausible to regard the construct as a composite of several traits.

Another related controversial issue raised by Rogers and Bateson (1991b) is whether test-wiseness is dependent or independent upon the subject area being assessed. This controversy, in fact, dates back to 1965 when Millman et al. published their article. On the one hand, Millman et al. (1965) noted that successful application of deductive reasoning strategies as well as most of cue-using strategies is dependent upon some partial knowledge of the subject matter being measured (pp. 713 - 714). The partial knowledge, although insufficient to know the correct answer directly, when combined with the test-wiseness strategies, could help an examinee figure out indirectly a possible correct answer. On the other hand, they stated in the addendum to their definition of test-wiseness that test-wiseness is independent of the examinee's knowledge of the subject matter being measured. The apparent inconsistency, according to Rogers and Yang (1996), may be attributed to Millman et al.'s intention in the addendum to reflect the generalization that test-wiseness is not restricted to a particular subject area. Hence, the

test-wiseness independence here should be interpreted as its applicability to a variety of subject areas and testing situations. While elements related to time using, error avoidance, and guessing are universally applicable to any testing situations, regardless of how much relevant knowledge the examinee has, the elements that require deductive reasoning or most of cue using (e.g., ID1, ID2, ID3, ID4, ID5, IIB2, and IIB5) are dependent on the relevant partial/minimal knowledge the examinee possesses (Rogers & Bateson, 1991b; Rogers & Yang, 1996). For example, to eliminate options which are known to be absurd (ID1), it is necessary to have minimal subject knowledge in order to know which options are wrong. Similarly, to determine whether two options are similar (ID2) or opposite (ID3), it is necessary to possess the partial relevant knowledge to tell whether they are similar or opposite. In short, partial knowledge is a crucial factor in the effective application of test-wiseness reasoning strategies.

2. A Systematic Variance of Measurement. Test-wiseness has been widely recognized as a systematic source of variance and, consequently, a potential threat to the valid interpretation of a test score (Thorndike, 1951; Millman et al., 1965; Sarnacki, 1979; Prell and Prell, 1986; Rogers and Bateson 1991a, b). Rogers and Bateson (1991a) suggested that an observed test score is a composite of an individual's knowledge of the subject area being assessed and 3 other components, two of which are partially determined by test-wiseness. In formula, it is expressed as:

$$X_{tot} = X_{kd} + X_g + X_{twd} + X_{er} > X_{kd}$$

where:

X_{tot} = the total number of correct responses,

X_{kd} = the number of correct knowledge derived responses and the score the test was initially desired to elicit,

X_g = the number of correct random responses,

X_{twd} = the number of correct test-wise derived responses, and

X_{er} = the number of correct "educated" random responses

(Rogers & Bateson, 1991a, p. 177).

Accordingly, the total scores on a test containing items susceptible to test-wiseness could be spuriously inflated by test-wiseness, incorporated with relevant partial knowledge that the examinee possesses.

3. Dimensionality of Test-wiseness. While validating his instrument to measure test-wiseness through the multitrait-multimethod approach (Campbell & Fiske, 1959), Millman (1966) incidentally found that the instrument consisting of four separate subscales (guessing, deductive reasoning, stem-option cue, and inconsistencies in the response options) did not quite hold up. The mean correlation of these subscales for monotrait-monomethod was .53 for the high school samples and .38 for the college students. For the college sample, some of the values in the main diagonal of the monotrait-monomethod submatrix were zero. One plausible explanation is that test-wiseness might not be unidimensional.

Likewise, examining the correlations among five subscales (longer correct alternatives, stem-option cues, specific determiners, grammatical cues, and similar

options), Diamond and Evans (1972) reported low correlations ranging from 0.02 to 0.33 among five subscales. Their findings suggested that use of the test-wiseness strategies was actually a set of several skills rather than a general skill.

Later, working with a sample of 520 fourth, fifth, and sixth-grade students, Benson (1985) employed first exploratory factor analysis and then confirmatory factor analysis to assess the dimensionality of test-wiseness. Benson obtained four factors which she tentatively labelled as (1) thoroughness, (2) preparation, (3) achievement motivation, and (4) perseverance. Correlations of -.08 to .57 among these factors indicated that, while correlated, the test taking skills involved (use of time, error avoidance, motivation, use of cues, and guessing strategies) were separable. Benson further argued that the alpha coefficient of the total test-wiseness was not high enough to verify the unidimensionality of test-wiseness.

In an attempt to convert Gibb's (1964) measure of test-wiseness into a shorter, more practical version of the test, Miller, Fuqua, and Fagley (1990) factor analysed the seven subtest scores obtained from a sample of 181 undergraduates enrolled in an educational psychology course. Using principal components followed by a varimax rotation, they found two factors. The first factor was tentatively labelled as "overt cues", involving stem-option cues, precise options, longer length option, and grammatical cues. The second factor was named as "subtle cues", including absurd options, specific determiners, and item giveaway (p. 207).

In another related study, Harmon, Morse, and Morse (1994) tested Miller et al.'s (1990) two-factor solution using confirmatory factor analysis approach. As a result, they reported that both a two-factor oblique model and a single-factor model fit the data

obtained from a sample of 173 undergraduates enrolled in educational psychology and speech pathology courses. Precise options, longer length options, and grammatical cues loaded on the first factor; absurd options and item giveaway loaded on the second. The correlation between the two factors was .72 (p. 13). Stem-option cues and specific determiners, however, did not load on any factor, regardless of whether it was in the two-factor solution or in single-factor solution. This may be attributable to the low mean scores (close to the chance level) for these two subtests.

In addition, Harmon et al. (1994) analysed the data of Miller et al. (1990) using confirmatory factor analysis. A two-factor oblique model rather than a single-factor model was confirmed. In terms of factor loadings, their two-factor solution was similar to what Miller et al. (1990) found, except for stem-option cues, which failed to load on either factor.

Based on Rogers and Yang's (1996) suggestion, the two factors found by Miller et al. (1990) and by Harman et al. (1994) may be interpreted, respectively, as the test-wiseness elements applicable in the absence of partial relevant knowledge and the cues that require the relevant partial knowledge. Nevertheless, taken as a whole, the results suggest that further study needs to be made on the dimensionality of test-wiseness.

In sum, although there is not sufficient evidence yet, research tends to support Benson's (1988) claim that test-wiseness appears to be a multidimensional construct, specific to item format, but general in relation to subject matter and persistent in nature (Millman, 1966; Diamond & Evans, 1972; Slakter et al., 1970; Woodley, 1973; Crehan et al., 1974; Sarnacki, 1979; Benson et al., 1986; Rogers & Bateson, 1991a).

Correlates of Test-wiseness

In an effort to validate test-wiseness as a construct, Millman et al. (1965) and Sarnacki (1979) suggested that the correlates of test-wiseness should be studied and that both convergent and divergent evidence should be collected using multitrait-multimethod procedures (Campbell & Fiske, 1959). If the predicted relationship between test-wiseness and other established constructs is found through logical and theoretical considerations and empirical investigations, then the evidence for the validity of the construct can be obtained. If not, further refinement of the construct, or of the measuring instruments, or both, is necessary. Following this suggestion, a large number of empirical investigations have been conducted to study test-wiseness in relation to such variables as intelligence, verbal ability, test anxiety, age, educational level, gender, race, and socioeconomic level.

1. Intelligence. Since test-wiseness is defined as a cognitive ability or a set of abilities (Sarnacki, 1979), it is logically expected that this construct should positively related to intelligence (Stanley, 1971). However, contrary to such an expectation, only weak to moderate correlations between test-wiseness and intelligence have been found. Ardiff (1965) conducted the first research in this area. When she administered her test-wiseness measure together with an intelligence test to 44 third graders and 48 sixth graders respectively, she found a correlation ($r = .51$) between the two instruments at the third grade but not at the sixth grade ($r = -.01$). Yet in another study, Diamond and Evans (1972), working with a sample of 95 sixth-grade students, reported moderate, positive correlations between intelligence, as measured by the Lorge-Thorndike Intelligence Test, and three specific test-wiseness cues----stem-cue, specific determiners and grammatical inconsistency----but not with two other cues----longer alternatives and similar options.

The variability of the correlations among various test-wiseness strategies and intelligence, as the authors argued, might indicate that test-wiseness is specific to certain cues only, and, therefore, not a general trait. More recently, in his analysis of the responses from 10,000 college applicants on the Scholastic Aptitude Test-Verbal (November 1984), Angoff (1989) found that success in guessing with partial knowledge was not only proportional to an examinee's ability but also dependent upon his or her personal idiosyncrasies----the willingness or reluctance to guess.

2. Verbal Ability. Another variable that is hypothesized to correlate positively with test-wiseness is verbal ability insomuch as identification of extraneous cues requires knowledge of grammar, vocabulary, and sentence structure. It is reasonable to expect that a test-wise individual would also possess high verbal ability. Empirical examinations have consistently supported this hypothesis. Using some common test-wiseness cues (e.g., stem-options, specific determiner, and similar options), but different samples from elementary and high school students to adults, Diamond and Evans (1972), Rowley (1974), and Bajtelsmit (1975) all found high positive correlations between their scales of test-wiseness and verbal achievement. Nevertheless, Ayrer, Diamond, Fishman, and Green (1976) reported low to moderate correlations (-.02 to .46) between verbal skills and test-wiseness within a sample of 76 fifth- and sixth-grade inner city children. Although it is possible that these lower correlations are accounted for by the homogeneous character of the sample which was not randomly selected, further study is necessary to assess the relationship between test-wiseness and verbal achievement among sub-groups of test takers such as ethnic/cultural groups and ESL students.

3. Test Anxiety. Logically, it might be expected that there would be a negative correlation between test-wiseness and test anxiety. An examinee must possess some amount of composure and know how to control test anxiety and nervousness in testing situations before she or he is able to identify and profit from test-wiseness cues in the test items. Conversely, a student without this composure may become too anxious to capitalize on test-wiseness. However, the research findings in this area are limited and conflicting. While Millman (1966) reported that there was no relationship between the two variables, Bajtelsmit (1977) observed a negative relation. In agreement with Millman, Rogers and Bateson (1991a), in their recent study of a sample of 936 Grade 12 students, found that test-wiseness reasoning was virtually uncorrelated with test anxiety; the correlations corrected for attenuation (Lord & Novick, 1968) were less than .17 ($_{c}r <$ .17) across the six subject area samples.

4. Grade Level and Gender. Given the complex nature of test-wiseness, it is logical to anticipate that test-wiseness would develop with increasing grade level and with test experience obtained from frequent practice and exposure to tests. Slakter et al. (1970) first studied test-wiseness in relation to sex and grade level. Using a test containing 16 test-wiseness items embedded within 28 regular items, they examined the development of four test-wiseness elements (i.e., stem-options, absurd options, similar options, and specific determiners) in a sample of 2,361 students from grades 5 through 11. They reported that both the reliabilities and the means of similar option and specific determiners subscales increased at higher grade levels. These increases, according to Slakter et al., suggested that there may be a developmental aspect of test-wiseness. In attempting to further validate the results from Slakter et al.'s study, Crehan, Koehler, and

Slakter (1974) conducted a longitudinal study by administering the same instruments to a sample of 1049 students in grades 5 through 11 who had been previously involved in Slakter et al.'s (1970) study. The authors of both related studies reached the same conclusions: although sex was not related to test-wiseness, grade level was. There was a steady increment in the acquisition of test-wiseness ability from grades 5 through 8, after which there was little or no further development. Sarnacki (1979), in agreement with these authors, suggested that "increased testing experience, maturation, and a general desire to achieve may aid high school students in attaining a common, asymptotic level of test-wiseness" (p. 270). His contention was later evidenced by Crehan, Gross, and Slakter (1978) in their second longitudinal study where a sex-by-year multivariate analysis variance revealed that test-wiseness increased with grade level over the period of 8 years and that large individual differences persisted into the high school grades. Furthermore, similar test-wiseness ability was found among college students (Gaier, 1962; Pryczak, 1973; Sax & Carr, 1962).

As a complement to research on the relationship between test-wiseness and age, Bajtelsmit (1975) scrutinized test-wiseness behaviour of adults and observed that adults possessed similar but deficient test-taking skills. Such deficiency, he suggested, might be attributable to their lack of recent exposure to tests.

5. Race and Ethnic Groups. Millman et al. (1965) postulated that test-wiseness found in objectively scored tests was culturally determined. Nevertheless, research findings on the relationship between test-wiseness and race, ethnic groups, or socioeconomic level are scattered and insufficient to arrive at any conclusive results. As

Scruggs and Lifson (1985) pinpointed, deficiency on test-wiseness skills exhibited by minority groups has simply been assumed rather than documented.

In an intent to validate Vernon's (1962) claim that the familiarity with particular item formats could not account for the score variance between British and American students (n = 183) on reading comprehension, Millman and Setijadi (1966) compared the performance of American and Indonesian students on three types of algebra items. Their findings showed that American students enjoyed an advantage on the multiple-choice questions, even after their Indonesian counterparts, who had no prior experience with this testing format before, were familiarized with the mechanics of choosing the correct answers.

Similar cultural differences were observed when Slakter (1969) compared test performance of Canadian and American Grade 8 students on Risk Taking on Objective Examination (RTOOE), a measured of risk-taking behaviour in the testing situation. Canadian students were outscored by their American peers. Obviously, this could not be attributed to language factor.

Lo and Slakter (1973) compared 131 Chinese twelfth graders in a city in Northern Taiwan with the same age-group students in the United States on the risk-taking behaviour and test-wiseness. They reported that Chinese students attained significantly lower mean scores on all four test-wiseness subscales (stem-options, absurd-options, similar-options, and specific determiners) than their American peers, but no differences were found in risk-taking behaviour as measured by RTOOE. In terms of test-wiseness, they observed that Chinese students tended to select, rather than eliminate, options with a specific determiner.

Similarly, Wu and Slakter (1978) compared risk-taking behaviour and test-wiseness of Grade 5, 8, and 11 Chinese students (n = 336) from both rural and urban schools with their American counterparts and found significant differences between the two national groups in terms of the relationships between grade levels and risk-taking behaviour measured by RTOOE. For Chinese students, risk-taking scores increased with the grade level, whereas, for American students, the opposite pattern was observed. The authors attributed such differences to the teacher's instruction on test preparation for highly competitive college entry examinations that Chinese high school seniors would write in near future. In addition, Wu and Slakter (1978) provided further support to the observation that Chinese students tended to select rather than eliminate options with specific determiners.

The findings of the two previous studies appear to evidence cultural differences in test-wiseness, and, subsequently, have been commonly cited. Although the results showed some differences between Chinese and American students in test-wiseness as well as in the application of test-taking strategies, the generalizability of these studies is questionable due to some methodological flaws. Inasmuch as the translation of all the instruments used in the studies from English into Chinese posed a major threat to the validity of the instruments themselves, any outcomes could be confounded by several underlying factors such as language differences, translation problems, test content with American cultural bias, or real differences in test-wiseness between two cultural groups. In this regard, further investigation on the peculiarity of the test-taking behaviours of Chinese students, as claimed by Lo and Slakter (1973) and by Wu and Slakter (1978), is needed.

Contrary to the findings of aforementioned studies, Diamond et al. (1976), in their study with black inner-city children, failed to find evidence to support the assumption that minority students lacked test-wiseness. Likewise, Urman (1983), in his study of 208 black, white, and Hispanic students in Grades 3 and 5, found that the interaction between race and test-wiseness training was not significant.

More recently, Man (1990) compared foreign Chinese students with Chinese immigrant students and Canadian students in terms of test-taking behaviours. While foreign Chinese students scored significantly lower ($p < .05$) than the other two groups in absurd-options, similar-options, stem-option cues and guessing subtests in the TTW, no differences were found between Chinese immigrant students and Canadian students in test-wiseness (Man, 1990). According to Man, the foreign Chinese student's deficiency in test-wiseness may be attributable to their inadequate English language proficiency. Thus whether ethnicity alone accounts for a significant amount of variance in test-wiseness is still a question.

To summarize, results of studies of test-wiseness suggest that test-wiseness is a complex construct, incorporating a number of behavioural and mental components. It appears to be (a) only weakly to moderately correlated with intelligence, (b) moderately to strongly correlated with verbal achievement, (c) possibly negatively related to test anxiety, (d) positively correlated with grade level but not with sex, and (e) possibly related to race and cultural background.

Susceptibility of Standardized Tests to Test-wiseness Cues

Teacher-made tests are commonly believed to be much more vulnerable to the influence of test-wiseness than standardized tests. Compared to the way standardized tests are constructed and the time devoted to this construction, most teachers have less time to develop their classroom tests. Further, they are less prepared to construct assessment instruments (Gullickson, 1986; Rogers, 1991). Consequently, they are less aware of test-wiseness principles and less sophisticated and precise in constructing items than the professionals involved in preparing standardized tests (Sarnacki, 1979; Mehrens & Lehmann, 1973; Benson, 1988; Rogers, 1991; Gullickson & Hopkins, 1987).

Nonetheless, when Ford and Weener (1980) attempted to empirically assess this belief with a sample of college freshmen, they unexpectedly found that not only teacher-made tests contained items susceptible to test-wiseness but also that standardized instruments possessed such items. And their findings were by no means accidental. According to Sarnacki's (1979) review and many other relevant studies (e.g., Metfessel & Sax, 1958; Bangert-Drowns, 1983; Benson, 1988; Slack & Porter, 1980; Smith, 1982; Rogers & Bateson, 1991a; Rogers & Wilson, 1993), standardized tests, though developed and/or reviewed through a vigorous process, are not necessarily immune to test-wiseness. For example, Metfessel and Sax (1958) reviewed 19 standardized measures, including aptitude, achievement, and personality tests. Based on their findings that the keying preference was located in the centre position in 29% of the cognitive instruments studied, and that 60% of the true-false items were true across all of the five personality tests involved, they concluded that test takers aware of these types of construction flaws would have more than average chance of scoring on a test. In Durost et al.'s (1970)

investigation and CTB/McGraw Hill's (1974) survey, two national standardized achievement tests, the Metropolitan Achievement Test (MAT) Form F, Level 5.0 - 6.9 (Durost et al., 1970) and the Comprehensive Test of Basic Skills (CTBS) Form S, Level 6.5 - 8.9 (CTB/ McGraw Hill, 1974), were analysed with the respect to six cues: (1) placement of correct answer, (2) length of the correct answer, (3) correct answer found in another item (item giveaway), (4) use of non-plausible distracters, (5) frequently use of "all" or "none of the above" as the correct answer, and (6) specific determiners and stem/option similarities. Both the MAT and CTBS were found to contain test-wiseness cues involving the placement and length of the correct answer. Twenty percent of the items in the CTBS and 6% of the items in the MAT were susceptible to item giveaway, i.e., one item providing the answer for another item. Similar findings that items are susceptible to test-wiseness have been reported in the examinations of other popular standardized instruments such as the Metropolitan Reading Test (Oakland, 1972), California Test of Basic Skills (Gains & Jongsma, 1974), Stanford Reading Test (Callenbach, 1973), and the Iowa Tests of Educational Development (Omvig, 1971). More recently, Rogers and Bateson (1991a) and Rogers and Wilson (1993) empirically assessed the influence of test-wiseness on the performance of high school seniors on two sets of provincial school leaving examinations in English, algebra, geography, history, social studies, biology, and chemistry. The susceptibility to test-wiseness found in those two sets of provincial examinations far exceeded what the test makers claimed about their standardized tests. For instance, in Rogers and Bateson's (1991a) study, they reported:

> The percentage of test-wise susceptible four-option multiple-choice items varied from 43% to 80% across the six examinations. The mean score of provincially representative samples on the subject area subtests comprised of these faulty items exceeded by 9.2% to 18.4% ($p < .01$) the mean score of the same students on the corresponding subtests consisting of the nonsusceptible items (p. 159).

The TOEFL is a standardized test containing items in multiple-choice format. Given the findings reviewed above, there is no apparent reason to accept any assertion that such a test is free from the impact of test-wiseness. Furthermore, despite non-existence of research on the effects of test-wiseness on the TOEFL, there has been a rapid growth in the number of TOEFL preparation programs. For example, claims are made in a brochure of one famous language school in China that, in their TOEFL preparation programs, approximately 70% to 80% of their students could successfully complete their courses within a period of 3 to 6 months and manage to reach or exceed a TOEFL score of 600, which is well above the cut-off scores required by most institutions in North America (Shanghai Progressing Institute, 1996). If there is any truth in their commercial advertisements, then the improved scores seem to depend more upon test-taking skills taught and practised within the TOEFL content than upon general English proficiency which usually takes much longer time than 3 or 6 months to develop (Des Brisay & Ready, 1991). In this case, students who are not exposed to such programs and training would inevitably be put in a disadvantageous position and subsequently penalized for the lack of relevant test-wiseness strategies. Accordingly, the following principles advocated by Standards for Educational and Psychological Testing (AERA, 1999) and by Principles for Fair Student Assessment Practices for Education in Canada (the Joint Advisory Committee, 1993) appear highly significant to this research:

> When test-taking strategies that are unrelated to the domain being measured are found to enhance or adversely affect test performance significantly, these strategies and their implications should be explained to all test takers before the test is administered. This may be done either in an information booklet or, if the explanation can be made briefly, along with the test directions. (AERA, 1999, p. 116).

> Interpretations of assessment results should take account of the backgrounds and learning experiences of the students.

> Assessment results should be interpreted in relation to a student's personal and social context. Among the factors to consider are age, ability, gender, language, motivation, opportunity to learn, self-esteem, socio-ecomonic background, special interests, special needs, and "test-taking" skills. Motivation to school tasks, language capability, or home environment can influence learning of the concepts assessed, for example. Poor reading ability, poorly developed psycho-motor or manipulative skills, lack of test-taking skills, anxiety, and low self-esteem can lead to lower scores. Poor performance in assessment may be attributable to a lack of opportunity to learn because required learning materials and supplies were not available, learning activities were not provided, or inadequate time was allowed for learning. When a student performs poorly, the possibility that one or more factors such as these might have interfered with a student's response or performance should be considered (Joint Advisory Committee, 1993, p.12).

Given the great influence of the TOEFL scores on the tertiary admission process and on professional certification in North America and elsewhere, and based upon the findings in the literature review presented above, the need was justified for the present study to investigate the influence of test-wiseness upon student performance on the TOEFL and to gain an understanding of the use of test-wiseness by Chinese examinees when responding to the TOEFL.

CHAPTER III

METHOD

As presented in Chapter I, the principal objective of the present study was to investigate the influence of test-wiseness upon performance on the TOEFL and the test taking behaviour of Chinese students when responding to a TOEFL item susceptible to test-wiseness. In order to achieve this objective, methods involving both quantitative and qualitative approaches were adopted in the study.

Described in this chapter are the instruments, samples, data collection procedures, and data analyses used in the research. First, a delineation of the instrumentation with relevant rationales and examples is provided. Next is a description of the samples, including both the initial sample and the interview sub-sample, and the way these samples were obtained. This is followed by a brief description of the two-stage procedure for data collection. The chapter concludes with a description of the data processing procedures, including the processes for scoring and data entry, and the statistical analysis and protocol analyses performed on the data gathered.

Instruments

Test of Test-wiseness

The Test of Test-wiseness (TTW) (Rogers & Bateson, 1991a) was adopted with some minor modifications made to suit the present study. To rule out the possibility of correctly responding to an item by content knowledge alone, the content of the items included in the TTW is either very unfamiliar to the participants or, in the case of some of

the incorrect options, very familiar. The modified TTW was designed to assess the following five test-wiseness elements listed in the Taxonomy outlined by Millman et al. (1965):

1. Three deductive reasoning strategies:

   ID1 – Eliminate options known to be incorrect.
   ID2 – Choose neither or both of two options which imply the correctness of each other.
   ID3 – Choose neither or one of two options, one of which, if correct, would imply the incorrectness of the other.

2. Two cue-using strategies

   IIB4 – Recognize and use similarities between the stem and the options.
   IIB3 – Recognize and make use of specific determiners.

The selection of the first four elements (i.e., ID1, ID2, ID3, and IIB4) was based upon the most frequently occurring test-wiseness cues identified in standardized tests involved in previous studies (Metfessel & Sax, 1958; Bangert-Drowns, 1983; Benson, 1988; Slack & Porter, 1980; Smith, 1982; Rogers & Bateson, 1991a). The specific determiner element (IIB3), a cue-using strategy, was added with the intention to further validate the findings of Lu and Slakter (1973) and Wu and Slakter (1978) that Chinese students tended to select rather than eliminate an option with a specific determiner. All five elements were selected to assess the extent to which the participants were able to use deductive reasoning and/or cue strategies to answer items unfamiliar to them in content but flawed in construction.

The modified TTW, hereafter simply referred to as the TTW, consisted of 26 multiple-choice items, the content of which was distributed approximately evenly across four subject areas–English, Mathematics, Social Studies, and Biology (see Table 3). All 26 items were selected from test-wiseness tests developed by Gibb (1964), Millman (1966), and Slakter et al. (1970). Twenty-two items were used and validated by Rogers and Bateson (1991a, b), one by Flippo (1987), and the remaining three items related to specific determiners by Weiten (1984). Both the reliability and validity of the items were proved to be acceptable (Rogers & Bateson, 1991a; Flippo, 1987; Weiten et al., 1980).

Table 3

Description of the Test of Test-wiseness

Test-wiseness Elements	Content Area				
	English	Math	Social stu.	Biology	Total
ID1	2	1	2	1	6
ID2	2	1	1	2	6
ID3	2	2	1	1	6
IIB4	1	1	1	2	5
IIB3	1		2		3
	–	–	–	–	–
Total	8	5	7	6	26

Note: ID1 - eliminate options known to be incorrect.
ID2 - choose neither or both of two options which imply the correctness of each other.
ID3 - choose neither or one (but not both) of two options, one of which, if correct, would imply the incorrectness of the other.
IIB4 - recognize and use similarities between stem and the options.
IIB3 - recognize and make use of specific determiners.

The testing time allocated was 30 minutes. The final version of the TTW is included in Appendix A. As shown in Appendix A, special instructions were given to encourage students to use test-wiseness skills, whenever possible, in responding to items on the TTW.

The scoring system employed to calculate the scores for the TTW was designed to reflect the use of the test-wiseness strategies considered. Modelled after Millman (1966) and Rogers and Bateson (1991a), the items were scored as follows:

One point for

ID1: correct answer;

ID2: either of the two non-similar options;

ID3: either of the two opposite options;

IB4: the cued option, and

IB3: correct answer.

Test of English as a Foreign Language

With the permission of ETS to use the two practice tests in the TOEFL Practice Tests (ETS, 1995d), the second of the two practice tests, the TOEFL Practice Test B, was employed in the research. Initially used on August 5, 1995 as a standard TOEFL, the Practice Test B consisted of Listening Comprehension (50 items), Structure and Written Expression (25 items), and Reading Comprehension (55 items). The allocated writing time was 1 hour and 55 minutes.

The Interview Form

The Interview Form was administered on an individual basis to further investigate the actual process each interviewee took when responding to each of the selected items. The Form included written instructions for the interviewees, followed by two subtests: (1) 13 items selected from the modified TTW (see pp. 58-60) initially administered to the full sample of the 390 Chinese students, and (2) 22 items selected from the TOEFL Practice Test B (ETS, 1995d). Using the results of the item analysis of the TTW based on the full sample (n = 390), 13 TTW were identified and selected across 5 test-wiseness elements (ID1, ID2, ID3, IIB3, and IIB4) and 4 subject areas (English, Social Studies, Mathematics, and Biology). There was clear evidence that, for those items, the students appeared to have employed test-wiseness skills.

Based on the author's judgement, 22 items were chosen from Section 1 (Listening Comprehension) and Section 3 (Reading Comprehension) of the TOEFL Practice Test B (ETS, 1995d), some of which were suspected of being susceptible to one of the ID1, ID2, ID3, IIB4 and ID4 test-wiseness elements. The nature of the selected items of the TOEFL was as follows:

Section 1: Listening Comprehension

(1) Items 10, 11, and 27 selected from Part A, Listening Comprehension, each of which involves a short conversation between a man and a woman, and followed by a question asked by a third person;

(2) Items 34-37 from Part B, Listening Comprehension, which is based on a long conversation between a man and a woman talking about the museum the woman visited, followed by four questions raised by a third person.

(3) Items 42-46 from Part C, Listening Comprehension, which is based on a mini-lecture on prehistoric desert people of Nevada, followed by 5 questions asked by another person.

Section 2: Reading Comprehension

(4) Items 41-50 excerpted from the last passage of the Reading Comprehension subtest. This item involved problems facing the United States after the Civil War.

A copy of the Interview Form is presented in Table 4 and the items included in are provided in Appendix B.

It should be pointed out that Section 2 (Structure and Written Expression) was not included in the present study. Given that the purpose of Section 2 is to measure the ability to identify grammatical errors from among the options provided, both recognition of grammatical errors and use of deductive reasoning to rule out erroneous options are intertwined inextricably, and considered construct-relevant. Therefore, it was not viable nor appropriate to detect presence of test-wiseness cues in Section 2. In addition, grammar has always been a major focus in Chinese ESL learning and, consequently, Chinese students in general feel more competent to deal with Section 2 than with other two sections such as Listening and Reading. When examinees possess enough knowledge to know the correct answer, application of test-wiseness skills related to cue-using and "educated guessing" strategy would be limited and often unnecessary.

Table 4

Description of the Interview Form

Content Area	Test-wiseness Elements							
	ID1	ID2	ID3	IIB4	IIB3	ID4	No test-wise cue	Total
The TTW **								
English	2	1	2	-	-	-	-	5
Mathematics	-	-	1	-	-	-	-	1
Social Studies	-	2	-	-	3	-	-	5
Biology	-	-	-	2	-	-	-	2
Total	2	3	3	2*	3	-	-	13
The TOEFL***								
Section 1: Part A	-	-	1	-	-	-	2	3
Part B	-	-	3	-	-	-	1	4
Part C	3	-	-	-	-	-	2	5
Section 2: Part A	-	-	-	-	-	-	-	-
Part B	-	-	-	-	-	-	-	-
Section3: Reading	4	1	1	-	-	1	3	10
Total	7	1	5	-	-	1	8	22

* One item used as warm-up practice

** Selection of TTW items based on the item analysis of the TTW

*** Selection of TOEFL items susceptible to various test-wiseness elements based on the author's judgement

Participants

Two samples of students from the Shanghai Qianjin Institute in Shanghai, China were selected. The initial sample was comprised of 390 students (average age of 22.8, 220 females and 170 males) enrolled in the 3-month advance TOEFL preparation program. They were typical potential TOEFL candidates in China: their mother tongue was Chinese (Mandarin); they had learned English since Grade 7; they either had already obtained or were close to obtaining a university degree in the areas of engineering, science, management and administration of international trade and business, and social studies and humanities; and they were seeking admission to graduate programs in colleges and universities in the United States and Canada. At the time the TTW and the TOEFL were administered, approximately 31% of these students had registered to write an official TOEFL within the next month. While the majority of them (97.3%) said that they had practised writing the TOEFL using previous forms or practice forms, none claimed to have been previously exposed to the particular instruments used in the research. All participants volunteered to be involved in the study and treated the TOEFL Practice Test B (ETS, 1995d) as a mock TOEFL test. As an incentive to participants, each participant was promised that she/he would be provided with an estimated TOEFL score shortly after the test administration.

The second or interview sample consisted of 21 females and 19 males selected from the initial sample of 390 students. Based upon the test scores of the TTW, two subgroups were selected: 10 female and 13 male highest scoring participants who scored 16 points or above, i.e., at least one standard deviation (SD = 2.97) above the mean ($\bar{X}$ = 13.26) on the TTW, and 11 female and 6 male lowest scoring participants who scored 10

points or below, i.e., at least one standard deviation below the mean on the TTW. They were asked to "think aloud" as they responded to the items in the Interview Form.

Data Collection Procedures

The data were collected in two stages. At Stage 1, the nature and objectives of the research were briefly explained to the initial sample of participants (n = 390). Following their consent, the TTW was first administered and then the TOEFL, with a ten-minute recess between the two tests. To enhance validity, the TOEFL test was administered strictly according to the rules and regulations established by ETS for a regular administration. The total testing time required was, respectively, 30 minutes for the TTW and 1 hour and 55 minutes for the TOEFL.

Immediately following the administration of the TTW and the TOEFL, the TTW was marked and the test scores were entered into a computer data file. Using the test scores of the TTW, the interview sample was identified.

Stage 2 of the data collection occurred approximately 10 days after the initial administration. The subsample of 40 participants selected according to their TTW scores (see previous section) were interviewed on an individual basis. Each participant was asked to "think aloud" about the strategies she/he was using while responding to each item. They were allowed to speak either in their first language (Chinese) or in their target language (English), or both. The interview was audio-taped. To minimize interviewee's anxiety and suspicion, field notes were not taken during the interview. The total time required for each interview was approximately one hour.

In each interview, the author met with a single student in a quiet room. Both of them sat side-by-side at a table on which there was a tape recorder and a folder containing the Interview Form. In an attempt to ensure the quality of the interviews, the following procedures were taken:

1. Prior to the beginning, approximately 5 minutes were set aside for informal conversation to establish rapport and to facilitate communication between the interviewer and interviewee;
2. Afterwards, the interviewer followed the script (see Appendix C for the details) that had been prepared to help each interviewee gain an understanding of the purpose of the study and the tasks to be undertaken. As a result, the interviewees were aware that the focus of the interview was on their test-taking behaviours and that their test-taking behaviours would not be evaluated in terms of appropriateness. To minimize any possible change of the test-taking behaviours, the interviewees were also told to take the test as they would under normal testing conditions;
3. In order to ensure that each interviewee was familiar with the think-aloud procedure, a sample item was first provided as an example and practised before they started working on the items; and
4. Interruption was kept to a minimum so that the situation would resemble a real testing situation. The interviewer remained as silent as possible while the interviewees were working on the items and explaining what they were thinking/doing. An effort was taken to ensure that question probes were used judiciously and would not provide clues to the interviewees about what to say.

In an effort to encourage the participants to report what they had actually done in processing the test rather than what they believe was expected of them, the rules of confidentiality and anonymity were pre-announced and strictly maintained throughout the research. The participants were explicitly told that their names and personal information would be kept confidential. Their individual test scores and recorded think-aloud protocols would be used anonymously for the research purposes only.

Data Analysis

Data Analysis for the TTW

The TTW was analyzed using LERTAP (Nelson, 1974), an item analysis computer program based on classical test theory. This program yielded the following statistics:

(a) At the item level --

The *p*-value, biserial, point-biserial, and mean for each option; and

(b) At the subtest level and test level --

The means, standard deviations, internal consistencies (Hoyt, 1949), and standard error of measurement.

Data Analysis for the TOEFL: Initial Sample

The identification of test-wise susceptible items in the TOEFL Practice Test B (ETS, 1995d) was determined in two stages: judgmental and empirical, to answer the first research question stated in Chapter I (p. 4):

Are there any test-wise susceptible items on the TOEFL? If yes, to what extent does the presence of such an item influence the test score and the interpretation of performance?

At the first stage (Judgmental), the researcher worked alone to identify which, if any, test-wiseness cues (ID1, ID2, ID3, IIB3, IIB4, and any other) existed in each item. Specifically, without listening to or reading the stem, an attempt was made to answer each question in Section 1 (Listening Comprehension) and Section 2 (Reading Comprehension) by eliminating options as incorrect and guessing the possible correct answer from among the remaining option(s). Those items which were correctly answered by "educated guessing" were considered tentatively to be susceptible to test-wiseness.

In an effort to establish the reliability of the findings, a peer checking technique (Ely et al., 1991) was applied. A volunteer with test development experience served as an independent judge. Following a brief introduction to the five target test-wiseness strategies (ID1, ID2, ID3, IIB3, and IIB4), the judge was asked to use the same approach to detect independently any test-wise susceptible item(s) from the same TOEFL test. As soon as this was done, a comparison was made and an agreement of 81% was found between the two separate sets of findings. Differences between the researcher and the judge were discussed to determine whether a consensus could be reached. Only the items that were identified as test-wise susceptible by both the researcher and the judge were considered to be susceptible to test-wiseness.

At the second stage (empirical), an item analysis of the responses of the initial sample (n = 390) was completed using the LERTAP program (Nelson, 1974). It was assumed that the *p*-values on the incorrect options would be approximately equal if each

of the distracters was equally plausible. Based upon the psychometric information (e.g., option *p*-values, point-biserials, and total test means), marked deviations from the distribution of *p*-values across the options and, for each option, the value of the point-biserial correlation coefficients with the total score and means on the TTW were examined along with a subject matter explanation of the results observed. Consequently, test-wise susceptible options in each item were identified and compared with the results from judgmental analysis. If the two findings were congruent, the presence of possible test-wise cues was confirmed.

Third, students' think-aloud protocols were analyzed and used to further validate the findings pertinent to the presence of possible test-wiseness cues in the TOEFL items.

With an intent to further reveal the influence of test-wiseness upon performance on the TOEFL, the correlations between the TTW and each item of the TOEFL were tested:

$$H_o : \rho_{pb} = 0$$

$$t = \frac{r_{pb}}{\sqrt{\frac{1 - r_{pb}^2}{n.-2}}}$$

where r_{pb} is the point-biserial correlation coefficient between the TTW and the TOEFL item, and n. is the number of the subjects in the sample (Glass & Hopkins, 1996, p.364).

Data Analysis for the Interview: Interview Sample

Protocol analysis (Ericsson & Simon, 1993) of the data obtained from the interview was conducted to answer the second and third research questions.

> What general cognitive processes do Chinese candidates usually employ when applying his or her test-wiseness, coupled with relevant partial knowledge, to respond to a test-wise susceptible item on the TOEFL?
>
> Is Rogers and Bateson's (1991a) model of test-wise test taking behaviour (see p. 3 for reference) for high school seniors generalizable to Chinese population? If not, what is distinctive about the group Chinese candidates who write the TOEFL in terms of their test-taking behaviour?

All recorded interviews were transcribed and then translated by the researcher. The transcriptions and translation were cross-checked for accuracy by another Chinese volunteer who had translation background. Discrepancies between the two translators were discussed until consent was reached.

In the analysis, attention was focussed upon the solution strategies employed by the test-wise subgroup and test-naive subgroup, including the test-wiseness strategies used and the sequence followed to produce answers in responding to each item in the Interview Form.

Based on Millman et al.'s (1965) Taxonomy of Test-wiseness Principles, a coding scheme was developed to track the test-wiseness strategies used by the students. Each student's specific test-taking strategies were scrutinized and compared to the classification of test-wiseness elements listed in the Millman et al. taxonomy. A new code plus a descriptive statement was generated each time a test-wiseness strategy distinct

from those in the taxonomy was identified. The frequency of each type of strategy was calculated.

In order to ensure reliability, another volunteer with a similar psychometric measurement background independently coded a random sample of 20 participants' interview think-aloud protocols. The codes she assigned were then compared with the researcher's codes. An agreement of 87% was found, with a majority of discrepancies occurring in coding the TOEFL subtest. With such a high agreement, the coding could be considered as reliable (Krippendorff, 1980).

CHAPTER IV

RESULTS AND DISCUSSION

TEST OF TEST-WISENESS

In this chapter, the test-wiseness abilities of Chinese students who were preparing to take the TOEFL are discussed. These abilities were assessed using the Test of Test-wiseness (TTW). The discussion starts with the test-wise behaviors of the entire sample (n = 390) on each of the subtests and on the total test of the TTW. This discussion is followed by the findings from the analysis of the "think aloud" data collected from the interview sample (n = 40) for the TTW items included in the Interview Form. The corresponding results for the TOEFL are presented in Chapter V.

Group Performance on the TTW

The mean, standard deviation, internal consistency, and standard error of measurement for the total sample of 390 respondents for each of the test-wiseness subtests and the total test are reported in Table 5. A copy of the results of the item analysis for the full sample is provided in Appendix D.

Table 5

Test Results of the TTW

TW Element	# of items	Mean (raw score)	SD[1]	Mean *p*-value (%)	% above chance	Internal Consistency[2]	SEM[3]
ID1	6	3.02	1.15	50.3	69.23	.10	1.0
ID2	6	4.01	1.21	66.8	87.69	.21	.98
ID3	6	3.13	1.23	52.1	70.00	.04	1.10
IIB3	3	.88	.79	29.3	21.28	.08	.62
IIB4	5	2.23	1.20	44.6	41.03	.24	.94
Total	26	13.26	2.97	51.0	80.77	.37	2.31

1 Standard Deviation
2 Hoyt Estimate (1941)
3 Standard Error of Measurement

ID1 Subtest

The mean for the ID1 subtest was relatively low ($\overline{X} = 3.02$) but beyond the upper bound of the chance score: i.e.,

$$\text{Upper Chance Score} = \frac{k}{a} + \sqrt{k \bullet \tfrac{1}{a} \bullet (1 - \tfrac{1}{a})}$$

$$= \frac{1}{4} \bullet 6 + \sqrt{6 \bullet \tfrac{1}{4} \bullet \tfrac{3}{4}}$$

$$= 2.56,$$

where k is the number of items and a is the number of options for each item.

Based on the score frequency distribution, 69.2% of the students scored above the chance level. Further data analysis at the item level revealed that Chinese students in general did well on items 3, 8, and 14, with the *p*-values for the correct option being 66.9%, 69.0% and 76.4%, respectively. However, quite a number of students (ranging from ⅔ to ¾) had difficulty in determining which was a correct answer to each of the three remaining items

(items 10, 18, and 23). For example, 31.0% and 37.7% of students were attracted, respectively, to Options C and D in item 10; 29.5% and 29.7% selected Options B and D, respectively, in item 18; and 35.4% chose Option C in item 23. The relatively low *p*-value for the correct answer for each of these three items resulted in the low mean for the ID1 subtest.

The internal consistency (Hoyt, 1941) was low (r = .10). This finding is comparable to the values reported by Slakter et al. (1970, r = .25) and by Rogers and Bateson (1990, r = .22). In contrast, the point-biserial coefficients were acceptable, ranging from .31 to .49 for all 6 items in the subtest. The apparent discrepancy between the low values for internal consistency and high values for point-biserials is likely attributed to guessing. Unless students are able to delete all three false options, they often have to guess from among the options remaining after elimination of options known to be incorrect.

ID2 Subtest

The mean for the ID2 subtest was 4.01, which exceeded the upper chance level ($X_c = 2.56$). The score frequency distributions showed that 87.7% of the students scored above the chance level. Item analysis revealed that the majority of students were able to successfully eliminate the two options containing similar elements for all items in the ID2 subtest except item 5, where nearly half (48.5%) of the students ($\bar{X}_{total} = 12.90$) selected Option D, one of the two similar options. Further analysis revealed that 39.5% of the well-performing students ($\bar{X}_{total} \geq 13.51$) chose either B or C and stayed away from the similar options in items.

The internal consistency (Hoyt, 1941), which was low (r = .21), is again consistent with the findings by Rogers and Bateson (1990, r = .36). Also, the point-biserial coefficients, ranging from .18 to .33, were a little lower than those found for the ID1 items. The low internal consistency and low point-biserial coefficients are possibly attributed to the double keying of these items in contrast to the single key for the ID1 items.

ID3 Subtest

The mean for the ID3 subtest was 3.13, exceeding the upper chance level (X_c = 2.56). Seventy percent of the students scored above the chance level. Data obtained from the item analysis showed that the students performing well on the TTW ($\bar{X}_{total} \geq 13.29$) tended to choose one of the two opposite options in all cases except item 11. For item 11, while 54.9% of the well-performing students picked one of the two opposite options, i.e., either A or B, 17.2% of them ($\bar{X}_{total} = 13.60$) were pulled to Option D. On the whole, the point-biserial coefficients were relatively acceptable, ranging from .22 to .38. Since the students had to guess between the two opposite options, the internal consistency (Hoyt, 1941) was low (r = .04). Once again, this low internal consistency is comparable to what was found about the opposite options (ID3) by Rogers and Bateson (1990, r = .16).

IIB3 Subtest

The mean (.88) of the IIB3 subtest was below the upper chance level (X_c = 1.50). Only 21.3% of the students scored above the chance level. The remaining students appeared to be insensitive to the cues of specific determiners (IIB3); their subtest scores were either at or below the chance level. The internal consistency (Hoyt, 1941) was also low (r = .08), which is exactly the same as the finding by Slakter et al. (1970, r = .08).

Further analysis of the data revealed that students who did choose the correct option for item 4 (38.7%), item 26 (35.4%), and item 9 (13.6%) also scored consistently higher on the TTW than those who failed to choose the correct option. The point-biserial coefficients for items 4, 9 and 26 were .67, .46, and .63, respectively. These high point-biserial coefficients once again confirmed the observation that only those students who possess strong test-wiseness skills have a better chance to score on the IIB3 subtest. Nevertheless, the students, including some strong ones, did not do well on item 9. While only 13.6% of the students successfully avoided the three options with specific determiners, 47.9% of the students selected Option C which did contain a specific determiner.

IIB4 Subtest

The mean for the IIB4 subtest was 2.23, essentially equal to the upper chance level (X_c=2.22). Examination of the subtest score distribution revealed that 41.0% of the students scored above the chance level.

The internal consistency was .24, which is exactly the same as the findings by Rogers and Bateson (1990, r = .24). The point-biserial coefficients were relatively high, ranging from .42 to .53. This suggested that the items containing stem-option connections discriminated well between test-wise and test-naïve students.

Summary of the Results for the TTW

The mean for the entire TTW was 13.26, which well exceeded the upper chance level (X_c=10.92). In addition, the score frequency distribution of the TTW showed that 80.8% of the students scored above the chance level for the TTW. These findings indicate that the majority of the Chinese students involved in this study possessed some test-wiseness skills. Further, since the percentages of the students scoring above the chance level for each of ID1, ID2, and ID3 subscales, were 69.2%, 87.7%, and 70.0%, it is likely that approximately 70% or more of the sample students were able to recognize incorrect options (ID1), similar options (ID2), and opposite options (ID3). The students, however, appeared to be less aware of stem-option connections (IB4). Only 41.0% of the students scored above the upper chance level on the IB4 subtest. Likewise, and somewhat more pronounced, the Chinese students seemed to be much less aware about how to handle options with specific determiners; 78.7% of the students scored at or below the chance level on this subscale. The latter finding is consistent with those of Lo and Slakter (1973) and Wu and Slakter (1978).

As expected from the previous research (e.g., Slakter, et al., 1970; Diamond & Evans, 1972; Allan, 1992; Rogers & Bateson, 1990), the internal consistency (Hoyt, 1941) was low for each of the subtests, ranging from .04 (opposite options) to .24 (stem-option cues), and also low (r = .37) for the entire TTW. However, the point-biserial coefficients for the majority (84.2%) of the items were acceptable, ranging from .20 (item 21) to .67 (item 4). As suggested by Rogers and Bateson (1991a), the low internal consistency may be attributed partly to the small number of items involved and partly to the student's "educated guessing" applied in response to a question which the students,

while able to eliminate one or two options as incorrect, did not know the answer from among the remaining options. When using "educated guessing", students tended to eliminate one or two options using their test-wiseness strategies, their partial knowledge, or a combination of both and then randomly guessed the answer from among the remaining options (Rogers & Bateson, 1991a). As shown next, the evidence that the students used educated guessing was found in the follow-up interviews.

Solution Strategies Used by the Interviewees when Responding to the TTW

In order to understand better what processes and strategies Chinese students used when responding to a test-wiseness susceptible item, a subsample of 40 students selected from the full sample of 390 students was interviewed. This subsample consisted of 23 "test-wise" ($X_{TTW} \geq 16$) and 17 "test-naïve" ($X_{TTW} \leq 10$) students. Following are the results and discussions of the analysis of the responses of these students.

Overall Performance of the Interviewees

The percentages of test-wise and test-naïve students who correctly responded to each item in the TTW of the Interview Form are provided in Table 6. As shown, the test-wise students outperformed their test-naïve peers in terms of the *p*-value for the correct answer for all items. For both the opposite options (ID3) and similar options (ID2) subtests, the test-wise students did better than the test-naïve students.

Table 6

Performance of Interview Samples on the TTW Interview Form

		Sample That Correctly Responded			
		Test-wise (n = 23)		Test-naive (n = 17)	
TW Cues	Item #	N	%	N	%
ID1	10	9	39.1	2	11.8
	23	11	47.8	8	47.1
ID2	5	14	60.9	4	23.5
	22	18	78.3	9	52.9
ID3	11	18	78.3	7	41.2
	21	12	52.2	3	17.7
	24	21	91.3	10	58.9
IIB3	4	15	65.2	9	52.9
	9	7	30.4	3	17.7
	26	10	43.5	7	41.2
IIB4	25	14	60.9	4	23.5
ID1, ID2 & ID3*	6	8	34.8	3	17.7
ID3	6	19	82.6	5	29.4

Note: * Only Option A considered to be the correct answer.

Why did the test-wise students perform better than their test-naïve counterparts? What test-wiseness strategies/skills did each subgroup of the students actually apply to make the difference in performance? To answer these questions, the 'think-aloud' protocols obtained from the interviews were analyzed. The results of this analysis are summarized in Table 7. The rows of the Table 7 correspond to items, clustered by the test-wiseness cues. The first column corresponds to the prime test-wiseness cues which in theory, if recognized and properly used, could lead to the correct answer. The second column contains the item number. Beginning with the third column on, the proportions of various strategies actually used by test-wise and test-naïve subgroups when responding to each TTW item are reported. Since the majority of students tended to employ more than one strategy in an effort to answer a TTW question, double or triple counting was often involved in calculating the proportion for each strategy applied. For example, item 10 contains absurd options (ID1) which could be eliminated using partial knowledge. However, it was found from the interview that, of the 23 test-wise students, 60.9% employed ID1, 47.8% used IIB4, 4.3% applied the convergence strategy, 4.3% relied on their knowledge, 17.4% attempted the answer using "educated guessing," and 13.0% picked the answer for its unusual length.

Table 7

Percentage of Test-Wiseness Strategies Used by Interview Samples When Responding to the TTW Interview Form

		SUBSAMPLE																			
		Test-wise (N = 23)										Test-naive (N = 17)									
TW	Item #	ID1	ID2	ID3	IIB3	IIB4	CV	PK	EdG	G	IIB1a	ID1	ID2	ID3	IIB3	IIB4	CV	PK	EdG	G	IIB1a	IIB1h
ID1	10	60.9	–	–	–	47.8	4.3	4.3	17.4	–	13.0	23.5	–	–	–	64.7	–	–	–	23.5	–	–
	23	87.0	–	–	–	–	4.3	–	21.7	8.7	–	58.8	–	–	–	–	–	–	5.9	41.2	–	–
ID2	5	43.5	69.6	8.7	–	21.7	–	–	69.6	4.3	–	11.8	23.5	5.9	–	41.2	–	–	23.5	11.8	–	17.6
	22	47.8	39.1	8.7	–	34.8	–	–	30.4	–	–	47.1	11.8	5.9	–	5.9	–	–	5.9	35.3	–	5.9
ID3	11	73.9	–	30.4	–	8.7	–	17.4	–	–	–	11.8	29.4	11.8	–	35.3	–	23.5	23.5	–	5.9	5.9
	21	34.8	17.4	47.8	–	8.7	–	–	43.4	–	–	58.8	17.6	–	–	11.8	5.9	–	5.9	23.5	–	5.9
	24	69.6	8.7	34.8	–	–	–	8.7	30.4	–	–	29.4	–	17.6	–	–	–	5.9	17.6	52.9	–	–
IIB3	4	26.1	–	–	65.2	–	8.7	4.3	8.7	–	–	35.5	–	–	29.4	–	11.8	17.6	11.8	17.6	5.9	–
	9	47.8	–	13.0	30.4	–	–	–	13.0	–	8.7	41.2	11.8	–	17.6	–	–	–	11.8	23.5	5.9	–
	26	52.2	–	–	60.9	–	8.7	–	17.4	–	4.3	52.9	–	–	29.4	–	–	–	17.6	11.8	11.8	11.8
IIB4	25	17.4	–	–	–	60.9	–	–	17.4	4.3	8.7	–	–	–	–	5.9	–	11.8	–	64.7	29.4	–
ID2&I	6	60.9	30.4	82.6	–	–	–	–	39.1	–	–	11.8	17.6	11.8	–	5.9	58.8	–	5.9	–	–	–

Note: Test-wiseness strategy counted regardless whether or not the final answer was correct or whether or not it was used alone or in combination with another.
ID1 = absurd options ID2 = similar options ID3 = opposite options IIB3 = specific determiners IIB4 = stem-option cues
CV = convergence PK = prior knowledge/common sense EdG =educated guessing IIB1a = longer/shorter options IIB1h=grammatical cues
G = guessing randomly, or intuition/hunch/gut feeling, or choosing particular option (e.g., C or D) or familiar/unfamiliar options

Performance on ID1: Eliminating Options Known to Be Incorrect

Item 10. **Wordsworth's "The Prelude" (1805)**
A. tells of a descent into Hell in a Model-T Ford.
*B. makes use of a distinction between "the sublime" and "the beautiful."
C. is concerned with the emerging African nations.
D. was influenced by Hemingway's **The Sun Also Rises**. (Rogers & Bateson, 1991a)

Item 23. **How many iambic feet (one iambic foot-one unstressed syllable followed by one stressed syllable, as in "perFORM") are in each line of Robert Pack's poem "The Compact"?**
A. 1
*B. 5
C. 16
D. 22 (Rogers & Bateson, 1991a)

Note: * correct option

Items 10 and 23 each contain three absurd options that should be eliminated using the partial knowledge. In theory, students who are able to identify and then eliminate at least one of the incorrect options will have a greater probability of identifying the correct answer.

Item 10. When answering item 10, 14 (60.9%) students in the test-wise subgroup used the ID1 strategy while the remaining nine (39.1%) students employed either the IIB4 (6 students) or IIB1a strategies (3 students). Of the 14 test-wise students who utilized the ID1 strategy, only five successfully recognized and then ruled out the three absurd alternatives. Two students also eliminated three options but one of the eliminated options was the correct answer. Both students thought the correct option too trivial to be true. Another student initially determined A and C as incorrect and then confirmed his judgment by citing what he believed to be a commonality between B and D, namely that both were associated with the stem in terms of literature. This commonality, according to

him, was more likely to be the place where the correct option was "hid." Similarly, three additional test-wise students first successfully eliminated two options as incorrect and then either randomly guessed between the remaining two options (1 student) or tried to identify the answer by establishing some link between the stem and one of the two remaining options (2 students). The remaining 3 of the 14 test-wise students first crossed out one option as incorrect and then selected C because, in their minds, the prefix "pre-" or the word "prelude" in the stem must be related to the word "emerging" in Option C. For example, one test-wise student explained:

> *I think that there might be relationship between "prelude" and "emerging". "Prelude" normally means an introduction to a piece of music. B is not likely because this is only about these two words. "The Prelude" looks like a piece of work. And piece of work cannot be just about the distinction between the two words. So, I eliminated B first. For A, I'm not sure. So, I chose between C and D. I guessed it must be C.*

In contrast, only four (23.5%) students in the test-naïve subgroup applied the ID1 strategy. The remaining 13 (76.5%) students chose the answer either utilizing the IIB4 strategy (9 students but none of them correct) or by randomly guessing (4 students). Of the four test-naïve students who used the ID1 strategy, only one successfully eliminated three incorrect options; one ruled out three options including the correct option; the remaining two first eliminated one or two options, and then picked the answer (both incorrect) using the IIB4 strategy.

Further, it seemed that more test-wise students went to Option C (8 students) than Option D (5 students) whereas more test-naïve students selected Option D (15 students) than Option C (4 students). The different test behaviors between the test-wise and test-naïve students appear to be attributable to the different amount of partial knowledge each

group students possessed. On the whole, the students in the interviews had very limited knowledge of African history and American literature due to their unique curriculum in China. This limited knowledge, while allowing the Chinese students, particularly test-wise students, to link the prefix "pre-" in the stem with the word "emerging" in Option C, was not enough to let them know when the African nations started to emerge. In contrast, for a large number of students, especially test-naïve students, their partial knowledge and limited test-wiseness skills were not sufficient to allow them to establish the lexical link between the stem and Option C. Instead, all they were able to do was to connect the stem with Option D based on their knowledge that both Wordsworth and Hemingway were writers. They did not know, however, that Hemingway lived after Wordsworth.

Item 23. Of the 23 test-wise students, 20 (86.9%) employed the ID1 strategy, two (8.7%) went with their "gut feeling," and one (5.9%) picked D because, according to him, 22 was the sum of the values for Options A, B, and C together (convergence strategy). Of the 20 students who used ID1, only seven successfully identified and then eliminated the 3 foils; three ruled out 3 options including the correct option; five eliminated one or two options as incorrect and then guessed among the remaining three or two options (3 correct); four misunderstood the stem and, consequently, crossed out 3 options including the correct answer; and the last one also misread the stem but got the correct answer for the wrong reason.

In contrast, 10 (58.8%) of the 17 test-naïve students used the ID1 strategy and seven (41.2%) guessed randomly. Of the 10 students who used ID1, three successfully recognized and eliminated the 3 distracters, two eliminated 3 options including the

correct option; five misunderstood the stem and, consequently, crossed out 3 alternatives including the correct options; and one misunderstood the question but got the correct answer for the same wrong reason as the test-wise student did.

The stem of item 23 seemed to be difficult for the Chinese students to understand. Four test-wise and five test-naïve students selected Option A simply because they misunderstood the question. They thought that they were being asked how many iambic feet there were in the word "compact" rather than in the poem "The Compact". Another two students (one test-wise and one test-naïve) took a similar approach by counting how many letters in the word "compact," and then, unable to find 7, took 5, the next closest number, from among the four numbers provided. For them, they selected the correct answer but for the wrong reason.

Eight students (4 test-wise and 4 test-naïve) selected Option C. Among them, one test-wise student claimed that he picked C out of his "gut feeling" and 3 test-naïve students chose C by randomly guessing. The remaining four (3 test-wise and 1 test-naïve) all claimed that they used reasoning. One student, classified as test-wise, stated:

> *I noticed that the numbers here are quite different. Since it is a poem, it should not be too short. One and five [iambic feet] are not appropriate. Twenty-two [iambic feet] seems too long. I once heard about Shakespearean sonnet, which is made of 14 lines. Here is 16, close enough. So, I guessed that it must be C.*

Her rationale was representative among those students who purposely picked Option C. It seemed that those students with their limited knowledge of English poetry confused iambic feet with the lines of a poem.

To summarize the differences in test-taking strategies used between test-wise and test-naïve students in the interviews when responding to the items with the ID1 cues, the

test-wise students appeared to be more capable than their test-naïve peers in recognizing and eliminating the absurd options using their partial knowledge. Also, compared with the test-naïve students, test-wise students not only more frequently but also more effectively employed the ID1 strategy. Their effective application of ID1 seemed to be attributable to their relatively wider knowledge domain. For instance, in item 10, 10 of the 23 test-wise students knew that Wordsworth could not be influenced by Hemingway because the former died long before the latter was born whereas only 2 of the 17 test-naïve students possessed this knowledge. This finding is consistent with Rogers and Bateson's (1991b) observation that test-wise and test-naïve students differ in knowledge. In addition, the test-wise students outperformed the test-naïve students in terms of effort/persistence in searching for cues. For example, test-wise students seldom guessed randomly from among the four options before exhausting their means/wisdom to look for cues.

Performance on ID2: Recognizing and Eliminating Similar Options

Item 5. Mr. Adams, in Henry Fielding's Joseph Andrews,

A. learns his parents were of the nobility.
*B. takes sick after falling through the ice.
*C. falls into the mud while reading.
D. discovers he is of noble birth. (Rogers & Bateson, 1991a)

Item 22. The treaty of Brest Litovsk was ratified by Moscow because

*A. Tsar Alexander I wanted to prevent Napoleon's invasion of Russia.
B. Russia was unable to keep up with the armament manufacture of Austria.
C. Russia could not keep pace with the military production of Austria.
*D. Nicolai Lenin wanted to get the Soviet Union out of World War I.
(Rogers & Bateson, 1991a)

Note: * the keyed options

Items 5 and 22 each contained a pair of similar options that, given there is only one correct answer, should be both eliminated. Students who were able to identify and rule out the two similar options received credit for choosing either of the remaining options. Overall, comparison of columns 4 and 14 in Table 7 reveals that students in the test-wise subgroup outperformed their test-naïve peers on these two items.

Item 5. The students in the test-wise subgroup well outperformed their test-naïve peers in item 5. Of the 23 test-wise students, 14 (60.9%) identified A and D as similar and eliminated both, and two (8.7%) considered A and D to be similar but not the same and then chose between these two options. Three (13.0%) students first eliminated two options as absurd, one of which was the correct answer, and then: one picked D, thinking the word "birth" as a cue linking the stem, and the other two simply guessed between the remaining two options (all incorrectly). Three (13.0%) test-wise students selected A or D because they thought that the stem plus Option A or D told a very dramatic story about a gentleman, who was "very poor at the beginning and later accidentally discovered that he was a noble descendent." Lastly, one (4.3%) test-wise student simply guessed.

Three (17.6%) of the 17 test-naïve students recognized A and D as similar and avoided both (ID2). One (5.9%) student classified B and C as the same category, first tempted to choose between B and C and then ruled out both because she could not tell which one was more likely to be the answer. One (5.9%) student appeared to use the ID3 strategy. She chose between rather than eliminated A and D, claiming that, while A and D were similar, they were not the same. Of the remaining 12 (70.6%) test-naïve students, two applied the ID1 strategy to eliminate three options as incorrect, one of which was the correct option; five chose A (1 student) or D (4 students), believing that the stem

combined with A or with D made "a nice dramatic story" (IIB4); three selected A or D, claiming that the selected option read better grammatically (IIB1h) and the last two simply guessed the answer.

Eighteen (45%) out of 40 students (8 test-wise and 11 test-naïve) selected Option D. This number is similar to the *p*-value (48.5%) for Option D based on the entire sample (n = 390). Further examination of the interview protocols showed that 15 out of these students picked Option D because they believed that the stem combined with Option D made a dramatic story (IIB4). This finding may shed some light into understanding why nearly half of the students in the full sample were attracted to Option D.

Item 22. This item seemed to be less difficult for most Chinese students. The *p*-value of the correct answers (double keys) for the entire sample (n = 390) was .66; for the interview sample (n = 40), the corresponding *p*-value was .68 (18 test-wise and 9 test-naïve). Yet it was found from the protocol analysis that the solution strategies actually adopted by the majority students in both the test-wise and test-naïve subgroups differed somewhat from what was expected. For example, of the 23 test-wise students, only nine (39.1%) identified and eliminated the two similar options (B and C) as expected, and then either employed the IIB4 strategy to select the answer (2), used the ID1 strategy to rule out the third option (5), or randomly guessed from the remaining two (2). Another six (26.1%) test-wise students picked Option D because, as they explained, Moscow did not become the capital until Lenin's era and, therefore, must be associated with Lenin/Soviet Union (IIB4). Likewise, another three (13.0%) test-wise students eliminated Options B, C, and D and then selected Option A because they believed that Lenin had nothing to do with the treaty of Brest Litovsk nor Russia with Austria (ID1). Among the remaining five

(21.7%) test-wise students, two chose between B and C, since they thought the answer must be between these two similar-but-not-the-same options (ID3), and three eliminated A and D as incorrect (ID1) and then selected randomly between B and C.

Of the 17 test-naïve students, two (11.8%) identified and then eliminated B and C as similar options (ID2), eight (47.1%) used their partial knowledge to eliminate three options as incorrect (6 correct and 2 wrong), one (5.9%) decided to choose between B and C because it seemed to her that B and C were opposite (ID3), and six (35.3%) simply guessed. It is worth mentioning that three of the eight test-naïve students who eliminated options using their partial knowledge deleted Options C and B for the wrong reason. When asked why B and C were incorrect, they all claimed that historically Russia had never been involved in the war with Australia [Austria]. Obviously, they misread "Austria" as "Australia". Another test-naïve student picked Option D simply because she strongly believed that only a good leader like Lenin wanted to get his nation out of the war.

To summarize the difference in the use of similar-option cues between test-wise and test-naïve students, the test-wise students were more successful in identifying and eliminating the similar options. The test-naïve students had difficulty not only in recognizing the similar options but also in appropriately employing the ID2 strategy. For example, a test-naïve student, while able to determine which options were similar, chose between them. It seemed that test-naïve students were also weaker readers. They misread the word 'Austria' as 'Australia' and, consequently, selected the correct option on item 22 but for the wrong reason.

Performance on ID3: Choosing between Opposite Options

Item 11. The literature of the early eighteenth century is

*A. public in nature, relating to society's outlooks and values.
*B. private in nature, relating to an individual's emotions and feelings.
C. rough and irregular compared to the literature of the later eighteen century.
D. filled with despair over the apparent collapse of traditional values.
(Rogers & Bateson, 1991a)

Item 21. "Lucifer in Starlight" is a

A. modern psychological story of World War II.
*B. Shakespearean sonnet by Gerard Manley Hopkins.
C. controversial French novel written by Resnais.
*D. Petrarchan sonnet by George Meredith. (Rogers & Bateson, 1991a)

Item 24. What is the probability that a needle of length L<D, when dropped on a table ruled with equidistant parallel lines of distance D apart, will cross one of the lines?

A. $\frac{\sqrt{3}}{3}(LD)$ *B. $\frac{2}{\pi D}$

C. L + .001D *D. $\frac{2D}{\pi L}$

(Rogers & Bateson, 1991a)

Note: * the keyed options

Items 11, 21, and 24 each contain a pair of opposite options, one of which is a correct answer. If one of the opposite options was chosen as the answer, then students would get one credit.

Based on the results of analysis of the interview data, students in the test-wise subgroup were able to make greater use of the opposite-option cue and/or other test-wiseness cues embedded in the ID3 items than their test-naïve peers.

Item 11. Of the 23 test-wise students, seven (30.4%) recognized that A and B were opposite and then chose one of these two options either randomly (4 students) or using their partial knowledge (3 students). Two (8.6%) chose C, claiming that the word "late" in the option made a good contrast to the word "early" in the stem (IIB4). The remaining 14 (60.9%) students relied on their personal knowledge or "common sense" and either eliminated incorrect options and then randomly chose from between the remaining options (4 out of 6 correct) or picked an option that made sense to them (7 out of 8 correct).

Although six (35.3%) test-naïve students were able to recognize A and B as opposite, four of them stayed away from these options rather than chose between the two. Another test-naïve student eliminated A and B as two similar options and then chose C due to the word "late" in the option in contrast to the word "early" in the stem. Five (29.4%) test-naïve students simply picked C based on the late-early connection (IIB4). One test-naïve student picked A because it was the shortest option (IIB1a). The remaining four test-naïve students picked B using their common sense (all correct).

It is noted that a large number of students (11 test-wise and 4 test-naïve) in the interviews chose B using their personal knowledge or "common sense". When asked why Option B made sense, one test-wise student explained:

> *I know that 18th century was the revolution era in the West. Individualism was a radical ideology against feudalism by then. Literature at that time naturally reflected the characteristics of this era, i.e., private in nature and relating to individual personal emotions and feelings. Therefore, the correct answer must be B.*

This student, while not quite familiar with the Western literature, knew something about Western history. Citing her learned rule that literature naturally reflects the

characteristics of the historical era, the student selected B. Her approach was representative of the interviewees who mainly relied on their partial knowledge for their answer to item 11.

Item 21. Eleven (47.8%) students in the test-wise subgroup recognized Options B and D as opposite and selected their answer randomly between the two options. While nine of these eleven students used the word "opposite" to describe B and D, two indicated that the options were similar because of the word "sonnet". Nevertheless, they did choose their answer from these two options rather than eliminated them. Four (17.4%) additional students in the test-wise subgroup also identified B and D as similar in structure. Nevertheless, they eliminated both options (ID2). The remaining eight (34.8%) test-wise students failed to recognize the ID3 cue and had to depend upon their partial knowledge. They either eliminated option(s) as absurd/irrelevant and then chose from the remaining option(s), or selected the answer which, they believed, might be connected to the stem. For example, one of these eight students described how she applied the ID1 strategy to arrived at her answer:

> *It was a hard decision mainly because I don't know what it means by "sonnet". It might mean a piece of music or a song. First, I don't think that C is correct because it [the stem] seems to have nothing to do with French. A is not likely, either. I don't feel that it is related to psychological analysis. B... "Shakespeare..." is not likely, either. I don't know why. I tended to choose D.*

Another test-wise student explained how she connected the stem with the option she chose as the correct answer:

> *I chose C because this [the stem] itself is unlikely to be a Shakespearean sonnet. Judging from its literary meaning, it has nothing to do with World War I. I guessed that it might be a novel. It sounds very much like a title*

for a novel. So, I chose C.

In contrast to the test-wise subgroup, none of 17 test-naïve students recognized B and D as opposite options. Instead, 10 (58.8%) relied on their partial knowledge to eliminate option(s) as absurd/irrelevant or to select the answer believed to be relevant to the stem. Four (23.5%) guessed randomly. The remaining three (17.6%) test-naïve identified B and D as similar and then crossed out both options. For instance, one test-naïve student explained why she eliminated B and D:

> *B and D are almost the same except for the different author followed. Both of them could not be the correct answer. C is awkward in structure. So, I chose A.*

Item 24. A similar pattern to what was found for item 11 was also identified from the student responses to item 24. Of the 23 test-wise students, eight (34.8%) recognized B and D as similar in structure but different in meaning and then selected one of them; and two (8.7%) eliminated B and D because both contained 'π'. The remaining 13 (56.5%) students failed to identify B and D as opposite. Of these 13 students, five had a problem understanding the question. Despite this, the 13 students all picked B, because they believed that, given $L < D$, and $\pi = 3.14 > 2$, only the value of the equation in Option B would be less than 1 under any circumstance. For example, one student explained how he got the answer this way:

> *I did not quite understand the question itself. I only understood the first part, i.e., "What is the probability that a needle of length L<D, when..." Sorry, I did not understand the rest of the sentence. However, I did know that the probability is less than 1. I asked myself which of the four options is less than 1. If L<D, then the answer must be B because* $\pi = 3.14 > 2$ *and* $\pi D > 2L$.

Among the 17 test-naïve students, three (17.6%) recognized B and D as opposite and selected one of them; five (29.4%) picked either B (3 students) or D (2 students) using their knowledge that the probability is less than 1, L < D, and $\pi = 3.14$. The remaining nine (52.9%) students simply guessed the answer.

Based on the student responses to the three questions containing opposite options, it seemed that identifying and guessing between the two opposite options was not always a prevailing test-taking skill of the Chinese students in the interviews. Instead, making use of resemblance between an option and the stem (IIB4) or a combination of IIB4 and ID1 was more commonly used. For the students who did recognize some relation between two opposite alternatives, there appeared to be a difference between test-wise and test-naïve students. When facing options with subtle differences, test-wise students were more willing to take risks, and, consequently, selected between the options, whereas test-naïve students were likely to avoid or eliminate both options even though sometimes they were aware that the two options were not exactly the same in meaning.

Performance on IIB3: Eliminating Items Containing Specific Determiners

Item 4. Hermann Klavemann is best known for

A. developing all musical scales used in the western world.
B. composing every sonata during the Romantic era.
C. translating all Russian classics into English.
*D. inventing the safety pin. (Rogers & Bateson, 1991a)

Item 9. In the Dartmouth College case the United States Supreme Court held that

A. the court had no right under any circumstances ever to nullify an Act of Congress.
*B. a state could not impair a contract.
C. all contracts must be agreeable to the state legislature.
D. all contracts must inevitably be certified.

(Weiten, Clery, & Bowbin, 1980)

Item 26. The Alabama claims were

A. all settled completely and satisfactorily.
B. claims against Jefferson Davis for seizure of all of the property in the state during wartime.
*C. claims of the United States against Great Britain.
D. claims of every citizen of Alabama against every citizen of Georgia.

(Weiten, Clery, & Bowbin, 1980)

Note: * the keyed options

Items 4, 9, and 26 each contain three options with specific determiners (e.g., all, every) that, if recognized, should be avoided or eliminated. If all the options with specific determiners are identified and ruled out, the correct option, which was the only alternative free of a specific determiner, will be found.

Item 4. In the test-wise subgroup, 15 (65.2%) students identified and eliminated all the options with specific determiners; two (8.7%) first ruled out D which, according to them, did not belong to the same semantic group as A, B and C, and then chose from the remaining options using their partial knowledge; one (4.3%) picked C because she believed that Hermann Klavemann was a translator whereas another five (21.7%) selected Option A, feeling that Hermann Klavemann was or read like a famous classical music composer.

In the test-naïve subgroup, five (29.4%) students recognized and ruled out all the options with the specific determiners; two (11.8%) first eliminated D which, they believed, did not belong to the same semantic group as Options A, B and C, and then chose from the remaining three options; six (35.3%) selected A because they believed that Hermann Klavemann was a well-known musician; one (5.9%) picked D because it was the shortest option (IIB1a); and three (17.6%) guessed randomly.

Item 9. While the test-wise students slightly outscored their test-naive counterparts on item 9, as a group they did not do well on this item. Of the 23 test-wise students, only seven (30.4%) successfully identified and eliminated the three options containing specific determiners. Three (13.0%) students guessed between C and D, claiming that the answer was likely one of these two options which were similar in structure but different in meaning (ID3); two (8.7%) students picked A for its unusual length (IIB1a); and the remaining eleven (47.8%) eliminated option(s) as incorrect and then selected an option as correct using their partial knowledge or "common sense"(1D1).

Three (17.7%) of the 17 test-naïve students successfully applied the IIB3 strategy to eliminate two or three options. Two (11.8%) students identified C and D as similar and then eliminated both options (ID2). One (5.9%) chose A because it was longer than the other three alternatives (IIB1a). Four (23.5%) randomly guessed. And the remaining seven (41.2%) ruled out options as incorrect and then picked the answer as correct based on their partial knowledge (ID1).

Of the 40 students who were interviewed, 13 test-wise and 9 test-naïve students altogether selected C. Twenty of these 22 students chose C because or partly because they felt that Option C made sense to them. When asked why Option C made sense to them, one test-wise student's answer exemplified the rationale for the students using the similar approach:

I thought that C is likely to be the answer because it is about the American Supreme Court's position on the conflict or law suit against a local organization. Since the America is a country governed by law and the Supreme Court is its highest law defender, it makes sense that they claimed that all contracts must be agreeable to the state legislature without any exception.

Item 26. Compared to the student performance on item 9, a similar pattern of solution strategies was found from student responses to item 26. The test-wise students once more did not do much better (p = 43.5%) on item 26 than their test-naïve peers (p = 41.2%) in terms of the proportion of students who chose the correct answer. However, the proportion of test-wise students able to recognize and make use of the IIB3 cues is higher than the proportion of test-naïve students. Fourteen (60.9%) of the 23 test-wise students recognized and made use of the IIB3 cues in responding to item 26: six successfully eliminated all three options with a specific determiner (IIB3); and eight ruled out two of the three foils containing specific determiners and then either randomly guessed (4 students) or used their partial knowledge (4 students) to choose from the remaining two options. In addition, one student picked B for its unusual length (IIB1a), and the remaining eight (34.8%) mainly relied on their partial knowledge either to eliminate options and then to guess from the remaining one(s) (3 students), or to select directly an option as correct or as relevant to the stem (5 students). Four out of the eight students who ruled out two options containing specific determiners successfully eliminated Options A and D using the IIB3 strategy. Yet when choosing between the remaining Options B and C, they found nothing wrong with B partly because they did not know who Jefferson Davis was and partly because, according to their own experience and cultural background with the violent and extreme communist revolution, they felt that it is not impossible for Jefferson Davis, regardless of who he was, to do something extreme like

seizing all of the property in the state during the wartime. For example, one test-wise student explained why she chose B:

> *Option A is impossible. How could claims be "all settled completely and satisfactorily?" The "dog" [D] is too hostile. C is likely in terms of its literary meaning. However, given the fact that it was the United States against Great Britain, I have never heard the Alabama claims. So, B looks more likely because such claims are likely to occur in the wartime.*

In contrast, five (29.4%) out of the 17 test-naïve students correctly utilized the IIB3 strategy to eliminate all three or two options with a specific determiner. Two (11.8%) test-naïve students selected A because it was shorter than any other alternatives (IIB1a), and two (11.8%) guessed randomly. The remaining seven (41.2%) test-naïve students used their partial knowledge to eliminate option(s) as incorrect and then chose from the remaining one as the correct answer (two correct).

In sum, a substantial difference in terms of ability to recognize and use IIB3 cues was found between the test-wise and test-naïve students. Nevertheless, test-wise students' potential was restricted by their unique cultural background. They tended to become insensitive to specific determiners particularly when the content of an item was related to violent, extreme, or non-mild events such as war, court suit, law making, claim, and declarations. This finding helps to understand why the mean score was relatively low (.88) on the IIB3 subtest for the entire sample (n = 390).

Performance on IIB4: Making Use of Stem-Option Cues

Item 25. The feulgen Nucleal Reaction demonstrates the presence of

* A. desoxyribonucleoprotein.
B. lysosomes.
C. mitochondria.
D. endoplasmic reticulum. (Rogers & Bateson, 1991a)

Note: * the keyed options

There exists lexical resemblance (i.e., the root 'nucle') between the key and essential part of the stem. If students are able to recognize and make use of this cue, they are more likely to correctly answer this item.

Item 25. Of the 23 test-wise students, 14 (60.9%) were able to recognize and link the root "nucle" in Option A with "Nucleal" in the stem. Four (17.4%) students eliminated Options B and C based on their personal knowledge and guessed between the two remaining alternatives (ID1). One (4.3%) student selected B because it was the shortest, and another one (4.3%) chose D because it contained two words (IIB1a). The remaining student (4.3%) simply guessed.

In contrast, only one (5.9%) student in the test-naïve subgroup identified and made use of the IIB4 cue. Of the remaining 16 test-naïve students, one (5.9%), whose subject area was biology, knew the answer; five (29.4%) picked the answer on the basis of whether it was shortest, longest, or contained two words (IIB1a); two (11.8%) chose the answer (wrong) based on their "hunch;" and the remaining nine (52.9%) randomly guessed.

Students in the test-wise subgroup did much better (p = 60.9%) on item 25 than their test-naïve counterparts (p = 23.5%) in terms of effective use of the IIB4 cue. Test-wise students tended to look for some subtle resemblance such as lexical/morphemic connections between the stem and the answer whereas test-naïve students tended to search for some obvious cues such as the shortest/longest option (IIB1a) or to give up and pick the answer randomly. In general, recognition and appropriate use of the IIB4 cue is a challenge for ESL students. For example in item 25, in order to recognize and use the lexical resemblance between the correct option and the stem, not only strong test-

wiseness skills are needed but also a sufficient knowledge on word formation is required.

Use of Multiple Cues

Item 6. The Proclamation of 1763

* A. forbade colonists to settle territory acquired in the French and Indian wars.
B. encouraged colonists to settle territory acquired in the French and Indian was.
C. provided financial incentives for settlement of territory acquired in the French and Indian war.
D. all of the above.

Note: * the keyed option

Item 6, a complex problem, requires synthesizing multi-steps/strategies. If a student can identify A and B as opposites, B and C as similar, and D as absurd because A and B are opposite, a unique answer, Option A, will be obtained.

Item 6. In the test-wise subgroup, eight students (34.8%) figured out the answer through synthesizing multiple cues (ID3, ID 2, and ID1). Eleven (47.8%) students did recognize A and B as opposites and chose between these two options whereas another four (17.4%) students identified B and C as similar, eliminated both, and then selected A.

In the test-naïve subgroup, two (11.8%) students selected A by recognizing and making use of multiple cues existing in the four options. One student (5.9%) identified A and B as opposite, but avoided both options. A fourth student picked A. According to her understanding, a proclamation was always made with an intention to forbid something. The remaining ten test-naïve students indicated that B and C were similar and, therefore, D was correct since D embraced at least two alternatives (ID4).

To summarize the comparison between test-wise and test-naïve students in synthesizing multiple cues, the test-wise students showed stronger ability. As a group, they enjoyed cognitive challenges and their strength lay in cognitive and metacognitive monitoring. Generally, they were able to look for cues from multi-dimensions and did not stop searching for cues until they were sure that there were no more cues or a satisfactory answer had been obtained. Although not all test-wise students in the interviews possessed the high-level test-wiseness skills required for arriving at the correct answer for item 6, they all identified at least one test-wise cue in the alternatives and capitalized on the cue identified. None of them simply chose D. When asked why not, one test-wise student explained that "the alternative of this kind was easiest to read but may be the least to be the answer." In contrast, the test-naïve students were less capable of looking for multiple cues. They tended to search for and be satisfied with superficial cues such as an option with the unusual length or the "all/none the above" option. Occasionally even when they did find a subtle or complex test-wise cue, they tended to avoid cognitive challenge/confrontation and did not capitalize on the cue identified. Instead, they tended to choose an option which was short in length and easy to read but least likely to be the correct answer. For example, ten test-naïve students, while able to tell that B and C were similar, did not bother to look further for more cues and simply went to D because they believed that Option D covered all, including the two similar options.

Summary of Chinese Students' Test-wiseness Behaviors Based on the TTW Interview Data

Based on the discussions above and on close examination of Table 7, three prominent issues have been summarized.

First, the difference between the test-wise and test-naïve students can be seen clearly in Table 7. In many cases, the prime strategies employed by the test-wise students were similar to the target strategies the TTW was designed to measure. Compared with the test-naïve students, the test-wise students' approaches seemed to be more thoughtful, logical, defendable, rational, and less random. They seldom selected an answer only because the option was the shortest/longest (IIB1a) or read better grammatically (IIB1h). Instead, they looked for more subtle and multiple cues. They tended to go for "educated guessing" whereas their test-naïve counterparts were likely to use random guessing. As a result, the test-wise students were more successful in figuring out the correct answer.

Second, based upon both the quantitative analysis of the entire group (n = 390) and qualitative analysis of the interview data collected from the interview group (n = 40), it is evident that Chinese students involved in the study did possess test-wiseness skills. Nevertheless, given their unique backgrounds and outlooks in culture, education, curricula, social values, ideology and political systems, the Chinese students, when searching for an answer, did not always employ the target strategies as expected in the development of the TTW. For example, for item 22, the prime strategy employed by the Chinese students was eliminating the absurd options using partial knowledge (ID1) rather than recognizing and then ruling out two similar options (ID2). Eleven test-wise and eight test-naïve students ruled out option(s) containing the word "Tsar" or "Russia" based

on their knowledge that Moscow did not become the capital until Lenin's era/the Soviet Union.

The ID1 strategy appeared to be the most frequently used strategy by both subgroups. Students in general tended to utilize this strategy either alone or in concert with other test-wiseness strategies. The reflecting mirror image of this approach was another consistently adopted strategy, i.e., finding stem-option connections using partial knowledge (IIB4). For example, in item 4, five test-wise and six test-naïve students chose Option A (wrong option) because they believed that Hermann Klavemann was or read like a famous classical music composer and thus should be connected with the word "musical" in Option A. Overall, it seemed to be a common approach among Chinese students to utilize their partial knowledge to eliminate options (ID1) or to identify a link between the stem and an option (IIB4). Their approach, nevertheless, was not always effective and successful with content that was unfamiliar to them. What is common knowledge in the North America and in the Western World may not be necessarily so for Chinese students or vice versa. For instance, when responding to item 10 which contained the ID1 cues, 14 test-wise and 4 test-naïve students attempted to figure out the correct answer using ID1 strategy. However, only 5 test-wise students and 1 test-naïve student successfully eliminated all three absurd alternatives. The remaining 12 students, while able to rule out one or two incorrect options, all failed to identified Option C or D as incorrect due to insufficient knowledge about when African nations emerged or when Hemingway lived. The lack of knowledge of African history and American literature may explain why 31% and 37.7% of 390 Chinese students involved in this study were, respectively, attracted to and selected Option C or D.

Incompetence in the English language may be another reason for the low score on some of the items. On item 23, for example, test-wise students performed poorly, as did their test-naïve counterparts. To a certain extent, this poor performance might be attributed to misunderstanding the stem. For example, four test-wise students chose Option A (incorrect) because they misunderstood the question as asking about how many stressed syllables there are in the word "compact" rather than in the poem "The Compact". Another four students (3 test-wise and 1 test-naïve) selected Option C because they did not understand the stem and confused iambic feet with the lines of a poem. This confusion could be one of the reasons that 35.4% of the 390 students involved in the study chose Option C in item 23.

Third, based on the item analysis of the entire sample of 390, the Chinese students overall did not do well on the three IIB3 items, especially items 9 and 26. In addition to possible insensitiveness to specific determiners, the lack of American cultural and historical background may be another factor contributing to the low scores on the IIB3 items. During the interviews, the majority (50% or more) of the test-wise and test-naïve students found nothing wrong with the specific determiners like 'all', 'every', and 'never' especially when those words appeared in laws, regulation, and legal/political claims during an extreme period such as wartime. In fact, such strong vocabulary does exist in the current national constitution and laws as well as many official documents in China.

In sum, Chinese sample students, particularly those classified as test-wise, did possess some test-wiseness skills including high level ones. In many cases, their approach and the way to deal with items containing test-wiseness cues was quite similar to the target strategies the TTW designed to measure. They also tended to eliminate or select option(s), or to establish the link between the stem and options using their personal knowledge. However, limited partial knowledge, incompetent language ability, and lack of cultural and historical background sometimes seriously jeopardized their effective use of these test-wiseness strategies.

CHAPTER V

RESULTS AND DISCUSSION

TEST OF ENGLISH AS A FOREIGN LANGUAGE

The results and discussion in the present chapter are focussed on the Chinese TOEFL candidates' test-wiseness behaviors in responding to the TOEFL Practice Test B (ETS, 1995d). The order of presentation is the same as that used in the previous chapter. First, the results of the analysis of the entire sample's (n = 390) performance on the TOEFL items susceptible to the ID1, ID2, ID3, ID4, and IIB4 test-wise elements are presented and discussed. This is then followed by findings from protocol analysis of the "think aloud" data from the interview subsample (n = 40) for the TOEFL items included in the Interview Form.

Overall Performance on the TOEFL

The mean, standard deviation, internal consistency (Hoyt, 1941), and standard error of measurement of the TOEFL, and correlation between the TOEFL and TTW are presented in Table 8. The mean is reported using the standard score metrics used by ETS (1995d) and the raw score, expressed as a percentage. The mean level of performance was 62.0%, 78.5% and 80.2% for the Listening Comprehension, Structure and Written Expression, and Reading Comprehension subtests, respectively. The corresponding standard deviations were 16.2%, 12.3%, and 15.0%, suggesting a fair amount of variability. The internal consistencies were, respectively, 0.87, 0.78, and 0.89.

The converted score means were 51 for Listening Comprehension, 57 for Structure and Written Expression, and 56 for Reading Comprehension. These values suggest that Chinese students involved in the study are knowledgeable of grammatical structure and their comprehension of written English is essentially the same and stronger than their ability to comprehend spoken English. This finding is consistent with what was reported in Yang's (1997) finding that Chinese ESL students in general are strong in grammar and relatively weak in listening and speaking.

Table 8

Overall TOEFL Results for the Entire Sample Students (n = 390)

Subtest	# of Items	Mean[1] (converted score)	Mean (% raw score)	SD (%)	Internal Consistency	SEM (%)	$r_{TOEFL/TTW}$[3]
Section 1: Listening Comprehension	50	51	62.0	16.2	.87	5.76	.32
Section2: Structure	40	57	78.5	12.3	.78	5.58	.37
Section 3:Reading Comprehension	50	56	80.2	15.0	.89	4.80	.39
Total	140	547	75.0	12.9	.84[2]	3.19	.40

[1] A TOEFL converted score is a scaled score transferred from a raw score. TOEFL section scores are reported on a scale of 20 to 68 for Sections 1 and 2 and 20 to 67 for Section 3; total scores are reported on a scale of 200 to 677. A total score of 600 or above is considered excellent whereas a score below 400 is regarded as weak.

[2] Cronbach's alpha for the composite

[3] Correlation between the TOEFL and the TTW

Turning to the total test score, the raw score mean was 75.0%, which corresponded to a converted score mean of 547. This converted TOEFL score was quite comparable to an average score of 532 for the entire population of the TOEFL examinees in China (ETS, 1997). Cronbach's alpha for the composite was 0.84, suggesting that a high degree of correlations among the three subtests. Based on the results of the subtests

and the total test, it appears that the sample of the Chinese students involved in the present study were typical of the TOEFL candidates in China.

The correlations between the TTW and Listening Comprehension, Structure and Written Expression, and Reading Comprehension subtests were, respectively, 0.32, 0.37, and 0.39. The correlation between the TTW and the entire TOEFL test was 0.40. These weakly moderate to moderate correlations are likely attenuated by the low internal consistency of the TTW (Rogers & Bateson, 1991a).

Presence of Test-wise Susceptible Items for the TOEFL

As discussed in Chapter III, items susceptible to test-wiseness in the TOEFL Practice Test B (ETS, 1995d) were first identified through consensus judgment done by two independent judges, and then verified by item analysis based on the entire sample (n = 390). Items were not considered as test-wise susceptible unless identified and verified by both the judgmental and empirical processes. Table 9 contains a breakdown of the types of test-wise cues found in each of the Listening and Reading Comprehension sections of the TOEFL. Section 2, Structure and Written Expression, was not considered in this study. Given that deductive reasoning is required in responding to Section 2, it is infeasible, if not impossible, to disentangle test-wiseness from the construct to be measured.

For the Listening Comprehension subtest, 24 (48.0%) of the 50 items were susceptible to the ID1 (k = 13) or ID3 (k = 11) strategy. In the case of the Reading Comprehension subtest, 32 (64.0%) of the 50 items were susceptible to various test-

wiseness including ID1 (k = 23), ID2 (k =1), ID3 (k = 1), IIB4 (k = 3), ID4 (k = 1), and ID5 (k = 3). To illustrate the rationale leading to each consensus judgement, six examples, one for each of the test-wiseness strategies listed in Table 9, are presented next.

Table 9

Presence of Test-wise Susceptible Items for the TOEFL

Subtest	No. of items	Test-wise Susceptible Items						Non-susceptible Items
		ID1	ID2	ID3	IIB4	ID4	Give-away (ID5)	
Section 1: Listening	50	13	-	11	-	-	-	26
Section 3: Reading	50	23	1	1	3	1	3	18
Total	100	36	1	12	3	1	3	44

Example 1: Items susceptible to Absurd Options (ID1)

Having read a passage regarding problems facing the United States after the Civil War, students were asked to respond to the following item (item 50, Reading Comprehension):

(50) It can be inferred from the passage that President Johnson pardoned the Southern Leaders in order to

	Section 3: Reading				Test-wiseness	
	n	p	r_{pbis}	$\overline{X}$	r_{pbis}	$\overline{X}$
(A) raise money for the North	2	.5	-.14	25.50	-.14	7.00
(B) repair the physical damage in the South	23	5.9	-.23	33.39	-.09	12.04
(C) prevent Northern leaders from punishing more Southerners	14	3.6	-.23	31.14	.02	13.57
*(D) help the nation recover from the war	351	90.0	.35	41.01	.09	13.27

Note: * the keyed option

On the right side of the options are the results of the item analysis for item 50 (n = 390). These results contain two components: (a) the psychometric data for each option of the item including the number of students who chose the option (n), the corresponding *p*-value (*p*), point-biserial (r_{pbis}), and mean ($\bar{X}$); and (b) corresponding point-biserial with the TOEFL total score and the TTW mean ($\bar{X}$) for the students who chose the option.

Option A in item 50 was identified as absurd by both judges. Based on the results of the item analysis, evidence of the application of ID1 is clearly seen. All the 390 students except two extreme test-naïve ($\bar{X}_{TTW} = 7.00$) avoided Option A, ostensibly because they realized that it did not make sense for the President to pardon Southern leaders in order to raise money.

Example 2: Items Susceptible to Similar-Option Cues (ID2)

It appears that The TOEFL Practice Test B (ETS, 1995d) was virtually free of items that contained similar-options (ID2). Across the two sections considered, only item 44 in Reading Comprehension was found to contain similar options. After reading the same passage identified above for item 50, the students were asked:

(44) According to the passage, which of the following statements about the damage in the South is correct?

	Section 3: Reading				Test-wiseness	
	n	*p*	R_{pbis}	$\bar{X}$	r_{pbis}	$\bar{X}$
*(A) It was worse than in the North.	291	74.6	.51	42.34	.20	13.56
(B) The cost was less than expected.	23	5.9	-.19	34.52	-.05	12.00
(C) It was centered in the border states.	47	12.1	-.34	33.32	-.13	12.48
(D) It was remedied rather quickly.	28	7.2	-.26	33.25	-.12	11.82
Omission	1	.3	-.03	36.00	-.03	11.00

Note: * the keyed option

Both judges reasoned that if "the cost was less than expected," it would be "remedied rather quickly." Based on the statistics obtained from the item analysis, it appears that the students tended to avoid these two options. As shown, the proportions of students who selected Option B (5.9%) or Option D (7.2%) were less than the proportion for the remaining incorrect Option C (12.1%). In addition, based on the TTW means, the students who chose Option B or D seemed to be weaker in test-wiseness than the students who selected the other options.

Example 3: Items Susceptible to Opposite-Options Cues (ID3)

The correct answer embedded between two opposite options was the second most common test-wise cue found in the TOEFL. Item 18 in Listening Comprehension was identified as containing opposite-option cues:

Students hear:

(Woman): Our football team didn't play well.
(Man): That's true, but at least we won the game.
(Narrator): What does the man mean?

Students read and then choose:	Section 1: Listening				Test-wiseness	
	n	p	r_{pbis}	$\overline{X}$	r_{pbis}	$\overline{X}$
*(A) The team won despite poor play.	293	75.1	.49	35.74	.18	13.51
(B) The team has to play at least one game.	13	3.3	-.17	26.15	-.09	11.62
(C) At least the football team played well.	27	6.9	-.27	25.41	-.15	11.41
(D) The team should have won the game.	57	14.6	-.31	27.35	-.06	12.68

Note: *the keyed option

In item 18, options A and D were identified as opposites, one of which was the correct answer. The item statistics show that the percentages of students who chose either of these two opposite options were greater than the percentages of students who

selected either of the two remaining options. The TTW mean for each of these two groups exceeded the corresponding means for each of the remaining groups. The response pattern for this item is typical of the outcomes of item analysis for an item susceptible to ID3.

Example 4: Items Susceptible to Stem-Option Similarity (IIB4)

Items were identified as susceptible to IIB4 in both the Listening and Reading Comprehension Sections. For example, in item 1 of Reading Comprehension, after a reading passage was presented, students responded to the following question:

(1) The author refers to the ocean bottom as a "frontier" in line 2 because it

	Section 3: Reading				Test-wiseness	
	n	p	r_{pbis}	$\overline{X}$	r_{pbis}	$\overline{X}$
(A) is not a popular area for scientific research	50	12.8	-.17	36.78	-.04	12.80
(B) contains a wide variety of life forms	9	2.3	-.33	23.89	-.16	9.78
(C) attracts courageous explorers	20	5.1	-.16	34.85	-.02	12.95
*(D) is an unknown territory	311	79.7	.36	41.47	.11	13.35

Note: * the keyed option

This item seemed to be designed to measure students' understanding of the main theme of the passage rather than their knowledge of the vocabulary word "frontier". Nonetheless, the two judges felt that Option D was likely to be the best answer. It contained a semantic connection with the word "frontier" in the stem because according to Webster Dictionary (1997), "frontier" literally means "an unknown territory" (Option D). Option A was a second possible answer, since, the word "frontier," by the semantic extension, means an unexplored field in scientific research (Option A). Option C was not completely absurd, for the word "explorers" in the option might be associated with "frontier". Options B was absurd because it was not semantically connected to "frontier"

at all. The results of the item analysis revealed that 79.7% of the 390 students did recognize the connections between the stem and Option D. The next most popular option was Option A, which was selected by 12.8% of the students. Only 9 test-naïve students ($\bar{X}_{TTW} = 9.78$) chose Option B.

Example 5: Items Susceptible to Convergence Strategy (ID4)

Only one item, item 41 in Reading Comprehension, was identified as susceptible to the convergence strategy, i.e., the correct option converged with or included other two or three options. Students were first presented with a reading passage about the situations after the Civil War, and then asked to choose the correct answer from among 4 alternatives.

(41) What does the passage mainly discussed?

Students read and then choose:	Section 3: Reading				Test-wiseness	
	n	p	r_{pbis}	$\bar{X}$	r_{pbis}	$\bar{X}$
(A) Wartime expenditures.	10	2.6	-.20	30.90	-.09	11.40
*(B) Problems facing the United States after the war	334	85.6	.43	41.44	.17	13.41
(C) Methods of repairing the damage caused by the war	32	8.2	-.23	34.38	-.05	12.59
(D) The results of government efforts to revive the economy	14	3.6	-.30	28.43	-.17	10.36

Note: * the keyed option

Without even reading the passage, both judges employed the convergence strategy to figure out independently that Option B was the correct answer because it encompassed Options C and D. In addition, Option A could be ruled out as absurd, since it did not belong to the same semantic group as the other three options. The statistics obtained from the item analysis support this consensus judgment: of the 390 students,

only 10 (2.6%) were attracted to A and their TTW mean score was low ($\bar{X}_{TTW} = 11.40$); 334 (85.6%) selected B, the correct answer, and their TTW mean score was highest ($\bar{X}_{TTW} = 13.41$); and 32 (8.2%) with relatively higher TTW mean score ($\bar{X}_{TTW} = 12.59$) chose C, which, while not completely wrong, was not as embracive as B.

Example 6: Items Susceptible to Give-away Cues (ID5)

The ID5 strategy involves using or finding the clues for the correct answer of one item in other items and/or options. This type of cue was found in items of the Reading Comprehension Section. For example, presented with a reading passage pertinent to population trends in postwar Canada, students were asked:

(17) When was the birth rate in Canada at its lowest postwar level?

	Section 3: Reading				Test-wiseness	
	n	p	r_{pbis}	$\bar{X}$	r_{pbis}	$\bar{X}$
*(A) 1966	376	96.4	.29	40.54	.11	13.25
(B) 1957	5	1.3	-.18	28.40	-.06	11.60
(C) 1956	4	1.0	-.20	25.50	-.11	9.50
(D) 1951	5	1.3	.12	32.20	-.02	12.60

(18) The author mentioned all of the following as causes of declines in population Growth after 1957 EXCEPT

	Section 3: Reading				Test-wiseness	
	n	p	r_{pbis}	$\bar{X}$	r_{pbis}	$\bar{X}$
(A) people being better educated	11	2.8	-.22	30.45	-.04	12.36
*(B) people getting married earlier	363	93.1	.38	40.91	.15	13.31
(C) better standards of living	6	1.5	-.12	32.83	-.10	10.50
(D) couples buying houses	10	2.6	-.29	26.70	-.11	11.00

Note: * the keyed option

The stem of item 18 suggests that population growth did not decline until after 1957. If this information is correct, then Options B, C, and D in item 17 could not be true. Based on this reasoning, without referencing the passage, it would be easy to rule out B, C, and D and to choose Option A as the correct answer for item 17.

Example 7: Items Susceptible to Multiple Test-Wiseness Cues

Items containing multiple test-wiseness cues were also identified in the TOEFL Practice Test B (ETS, 1995d). For example, item 8 of Listening Comprehension was found to contain ID3 and ID1 cues:

Students hear:

(Man): That's a long line! Do you think there'll be any tickets left?
(Woman): I doubt it – guess we'll wind up going to the second show.
(Narrator): What does the woman mean?

Students read and then choose:	Section 1: Listening				Test-wiseness	
	n	*p*	r_{pbis}	$\overline{X}$	r_{pbis}	$\overline{X}$
*(A) They'll have to go to a later show.	336	86.2	.32	34.50	.12	13.33
(B) The people in line all have tickets.	1	.3	-.08	21.00	-.06	9.00
(C) She doesn't want to go to the second show.	12	3.1	-.17	25.67	-.14	10.67
(D) They won't have to wait much longer.	41	10.5	-.25	27.66	-.04	12.78

Note: * the keyed option

In this item, options A and D are opposites, one of which is the correct answer. In addition, Option B seems to be absurd. When facing this item, students with strong test-wiseness ability tended to select between A and D, the two opposite options. Again, the TTW means for the groups who selected Option A or D are higher than for the remaining two options. Moreover, the students appeared to know that the conversation took place when the man and woman were waiting to buy tickets, and, consequently, ruled out Option B as incorrect (ID1). Of the 390 students, only one student picked B, and this student was weak in test-wiseness ($\overline{X}_{TTW} = 9.00$).

Summary of Test-wise Susceptible Items of the TOEFL

Table 10 contains a summary of group performance on the test-wise susceptible items and non-testwiseness items on the Listening and Reading Comprehension subtests of the TOEFL. Based on the results of correlated t-test, the means for the items identified as test-wise susceptible exceeded significantly ($p < .01$) the corresponding means for the non-testwiseness items in both the Listening and Reading Comprehension Sections. The standard deviations on the susceptible items and on the non-test-wiseness items were also significantly different at the .01 level for both sections considered. The correlation coefficients ($r_{ttw/toefl}$) between the TTW and each section of the TEOFL for test-wise susceptible items were moderately week, and not significantly different ($p < .25$) from the corresponding correlation coefficients ($r_{ttw/toefl}$) for non-testwiseness items. Again, the moderately weak correlations may be attributable to the low internal consistency of the TTW (see Table 6).

Table 10

Summary of Test-wiseness Analysis of the TOEFL (n =390)

Subtest	Test-wise Susceptible Items					Non-susceptible Items				
	# of Items	$\bar{X}$ (%)	SD (%)	R_{xx}	$r_{ttw/toefl}$	# of Items	$\bar{X}$ (%)	SD (%)	R_{xx}	$r_{ttw/toefl}$
Section 1: Listening	24	69.6*	17.5*	.77	.30	26	62.5	15.8	.74	.30
Section 3: Reading	32	80.4*	13.6*	.80	.38	18	76.1	18.5	.78	.34

Note: R_{xx} = Internal consistency (Hoyt, 1941).
$r_{ttw/toefl}$ = Correlation between the TTW and each subtest of the TOEFL.
* Significantly different from the corresponding mean on non-susceptible items ($p < .01$).

Solution Strategies Used by the Interviewees when Responding to the TOEFL

What was the major underlying cause for the higher means on the subsets of test-wise susceptible items? Did these items happen to be inherently easier? Or did these items become "construct-irrelevant easier" (Messick, 1989) when extraneous clues (i.e., test-wiseness elements) in items permitted students to respond in ways irrelevant to the construct being measured? Or were the high means attributable to the combination of both? Since any of these three reasons could be possible and valid, there is no ready statistical and psychometrical answer. To shed some light on these questions and understand what processes and strategies the Chinese students employed while responding to a test-wise susceptible item of the TOEFL, a subsample of 40 students was purposely selected from the full sample and asked to respond to the Interview Form of the TOEFL constructed for the present study. This subsample consisted of 23 "test-wise" ($X_{TTW} \geq 16$) and 17 "test-naïve" ($X_{TTW} \leq 10$) students. These students were the same students interviewed for TTW. During each interview, the students were presented with selected TOEFL items containing various specific categories of test-wiseness elements (see pp. 61-62) and asked to think aloud and/or describe how they got their answer or the reasons for their answers. The results of the analyses of their responses are presented and discussed next.

Overall Performance of the Interviewees on the TOEFL Interview Form

The percentages of students in the test-wise and test-naïve subsamples who correctly answered each of the 22 TOEFL items included in the Interview Form are reported in Table 11 together with the difference between the two percentages. As shown, the test-

wise students outperformed the test-naïve students by at least 10 percent on 15 items[1].

Table 11

Performance of the Interview Sample on the TOEFL Items of the Interview Form

		Sample That Correctly Responded				
		Test-wise (n = 23)		Test-naive (n = 17)		Diff
TW Cues	Item #	N	%	N	%	%
ID1	44 (Listening)	20	87.0	8	47.1	39.9
	45 (Listening)	23	100.0	15	88.2	11.8
	46 (Listening)	22	95.7	13	76.5	14.2
	42 (Reading)	17	73.9	5	29.4	44.5
	43 (Reading)	22	95.7	14	82.4	13.3
	47 (Reading)	22	95.7	16	94.1	1.6
	50 (Reading)	23	100.0	16	94.1	5.9
ID2	44 (Reading)	19	82.6	12	70.6	12.0
ID3	27 (Listening)	17	73.9	12	70.6	3.3
	34 (Listening)	21	91.3	15	88.2	3.1
	35 (Listening)	22	95.7	14	82.4	13.3
	37 (Listening)	19	82.6	11	64.7	17.9
	49 (Reading)	22	95.7	8	47.1	48.6
ID4	41 (Reading)	22	95.7	14	82.4	13.3
Non-testwise	10 (Listening)	23	100.0	14	82.4	17.6
	11 (Listening)	22	95.7	14	82.4	13.3
	36 (Listening)	16	69.6	5	29.4	40.2
	42 (Listening)	23	100.0	15	88.2	11.8
	43 (Listening)	20	87.0	13	76.5	10.5
	45 (Reading)	23	100.0	16	94.1	5.9
	46 (Reading)	23	100.0	17	100.0	0.0
	48 (Reading)	23	100.0	17	100.0	0.0

In contrast, the test-naïve students did not outperform the test-wise students on any of the items. There was no relationship between the differences and test-wiseness cues. For

[1] A statistical test was not conducted given its dependency on *p*-values. Instead, the difference of 10% was

example, the two subsamples performed essentially the same or exactly the same on two of the seven items containing an absurd option (ID1); on two of the five items containing opposite options (ID3); and on three of the eight items that were not susceptible to any test-wiseness elements in Millman et al.'s Taxonomy of Test-wiseness Principles (1965). The question then arose "why on some of items, both test-wise susceptible and non-test-wise susceptible, did the two groups perform essentially the same whereas on other items the test-wise students outperformed the test-naïve students?"

A summary of the solution strategies used by both test-wise and test-naïve groups when responding to the TOEFL items included in the Interview Form is presented in Table 12. The format of Table 12 is similar to that of Table 7 in Chapter IV. The rows of the table correspond to items, clustered by the test-wise cues. The first column corresponds to the test-wiseness cues identified by the two judges and confirmed by the results of the item analysis based on the entire sample (n = 390). The second column contains the item number. Columns 3 to 11 contain the proportions of the strategies the test-wise students actually used when responding to each TOEFL item. The corresponding proportions for the test-naïve students are provided in columns 13 to 23. The last column for each subgroup, i.e., columns 14 and 24, respectively, contains the proportion of students who claimed to know (K) the answer to each item. This proportion, however, is not necessarily consistent with the corresponding *p*-value for the correct answer, since what the students claimed to know may or may not have been the

---

considered to be large enough to influence test scores.

Table 12

Percentage of Test-Wiseness Strategies Used by Subgroups When Responding to the TOEFL

		SUBGROUPS																						
		Test-Wise (N = 23)												Test-Naive (N = 17)										
TW	Item #	ID1	ID2	ID3	ID4	T	IIB4	PK	EdG	G	PRE	O	K	ID1	ID2	ID3	ID4	T	IIB4	PK	EdG	G	PRE	O	K
ID1	44 (L)	39.1	-	-	-	-	60.9	17.4	4.3	-	95.6	-	-	29.4	-	-	-	-	47.1	-	-	11.8	82.4	-	-
	45 (L)	8.7	-	-	-	-	91.3	-	-	-	100	-	-	5.9	-	-	-	-	88.2	-	-	5.9	82.4	-	-
	46 (L)	26.1	-	-	-	-	47.8	-	-	-	87.0	-	26.1	23.5	-	-	-	-	23.5	-	-	17.6	70.6	-	35.5
	42 (R)	52.2	-	-	-	-	-	8.7	17.4	-	17.4	-	39.1	70.6	-	-	-	-	-	29.4	-	5.9	17.6	-	23.5
	43 (R)	8.7	-	4.3	-	-	-	-	-	-	17.4	-	71.4	11.8	-	-	-	-	5.9	11.8	-	-	11.8	-	70.6
	47 (R)	95.7	-	-	4.3	-	-	-	-	-	39.4	-	4.3	41.2	-	-	-	-	-	-	5.9	-	17.6	-	58.8
	50 (R)	95.7	-	-	-	-	-	-	-	-	21.7	-	4.3	47.1	-	-	-	-	-	-	-	-	17.6	-	52.9
ID2	44 (R)	8.7	-	-	-	-	8.7	17.4	4.3	-	39.4	-	65.2	5.9	-	-	-	-	-	23.5	5.9	5.9	23.5	-	64.7
ID3	27 (L)	34.8	-	30.4	-	26.1	17.4	34.8	17.4	4.3	100	-	4.3	11.8	-	11.8	-	17.6	-	58.8	11.8	11.8	88.2	-	-
	34 (L)	-	-	13.0	-	13.0	13.0	-	-	-	100	-	65.2	-	-	-	-	-	17.6	-	-	-	82.4	-	82.4
	35 (L)	21.7	-	13.0	-	-	-	34.8	-	4.3	100	-	60.9	17.6	-	-	-	-	-	17.6	-	-	82.4	-	82.4
	37 (L)	39.1	-	52.2	-	-	52.2	77.4	4.3	4.3	100	-	-	23.5	-	23.5	-	-	23.5	17.6	-	17.6	82.4	-	17.6
	49 (R)	26.1	-	8.7	4.3	-	56.5	-	-	-	17.4	-	8.7	11.8	-	-	-	-	23.5	-	5.9	5.9	17.6	-	64.7
ID4	41 (R)	47.8	-	-	56.5	-	-	-	-	-	71.4	-	43.5	17.6	-	-	-	-	-	-	-	-	29.4	-	82.4

Table 12 (Continued)

		SUBGROUPS																						
		Test-Wise (N = 23)												Test-Naïve (N = 17)										
TW	Item #	ID1	ID2	ID3	ID4	T	IIB4	PK	EdG	G	PRE	O	K	ID1	ID2	ID3	ID4	T	IIB4	PK	EdG	G	PRE	O	K
NON-TEST WISE	10 (L)	–	–	–	–	–	–	–	–	–	95.6	8.7	91.3	–	–	–	–	–	11.8	–	–	5.9	76.5	–	82.4
	11 (L)	13.0	–	–	–	4.3	4.3	–	–	–	100	–	82.6	–	–	–	–	–	–	–	–	–	76.5	–	100
	36 (L)	21.7	–	–	–	–	60.9	–	–	17.4	87.0	8.7	–	23.5	–	–	–	–	47.1	–	–	23.5	64.7	–	5.9
	42 (L)	–	–	–	–	–	100	–	–	–	100	–	–	–	–	–	–	–	82.4	–	–	–	82.4	–	17.6
	43 (L)	13.0	–	–	–	–	65.2	–	–	–	82.6	–	21.7	5.9	–	–	–	–	52.9	–	–	–	70.6	–	41.2
	45 (R)	4.3	–	–	–	–	–	8.7	–	–	21.7	–	87.0	–	–	–	–	–	–	–	–	–	17.6	–	94.1
	46 (R)	–	–	–	–	–	–	–	–	–	21.7	–	100	–	–	–	–	–	–	–	–	–	17.6	–	100
	48 (R)	–	–	–	–	–	–	–	–	–	13.0	–	100	–	–	–	–	–	–	–	–	–	17.6	–	100

Note: Test-wiseness strategy counted regardless whether or not the final answer was correct or whether or not it was used alone or in combination with another.
(L) = Listening Comprehension (R) = Reading Comprehension ID1 = absurd options ID2 = similar options ID3 = opposite options
ID4 = convergence T = the tone of the speaker IIB4 = stem-option cues PK = prior knowledge/common sense
EdG = educated guessing PRE = predict/read questions in advance O = other test-wiseness strategy K = know/claim to know the answer
G = guessing randomly, or intuition/hunch/gut feeling, or choosing particular option (e.g., C or D) or familiar/unfamiliar options

correct answer. Given that students often tended to use multiple strategies to solve a problem, double or triple counting is involved in calculating the proportion for each strategy. For example, item 34 in Listening Comprehension contains a pair of opposite options, one of which is the correct answer (ID3). It was found from the interview that 3 (13.0%) of the 23 test-wise students used multiple strategies. They first recognized A and C as opposites (ID3) and then selected Option A claiming that it contained the word they heard from the conversation (IIB4).

Looking at Tables 11 and 12, regardless of whether the *p*-value differences were large or small between the test-wise and test-naïve subgroups, in most cases there are discernible differences between the solution strategies each subgroup reportedly employed. These differences are described and compared next by test-wise strategy category.

Items Susceptible to ID1: Absurd Options

There were three listening and four reading items that each contained at least one absurd option. The results for the three listening items are presented and discussed first, followed by the results and discussion for the four reading items.

Items 44 to 46 in Listening Comprehension

Items 44 to 46 were selected from a set of five items presented following a short talk given by an anthropologist:

Students hear:

(Narrator): Questions 42 through 46. Listen to part of a talk given in an anthropology class.

(Man): Today's lecture will center on prehistoric people of the Nevada desert. Now, most of these prehistoric desert people moved across the countryside throughout the year. You might think that they were

wandering aimlessly – far from it! They actually followed a series of carefully planned moves. Where they moved depended on where food was available – places where plants were ripening or fish were spawning.

Now often when these people moved, they carried all their possessions on their backs, but if the journey was long, extra food and tools were sometimes stored in caves or beneath rocks. One of these caves is now an exciting archaeological site. Beyond its small opening is a huge underground grotto. Even though the cave's very large, it was certainly too dark and dusty for the travelers to live in – but it was a great place to hide things, and tremendous amounts of food supplies and artifacts have been found there. The food includes dried fish, seeds, and nuts. The artifacts include stone spear points and knives; the spear points are actually rather small. Here's a picture of some that were found. You can see their size in relation to the hands holding them.

(44) Why didn't people live in the cave described by the speaker?

Students read and then choose:

(A) They had trouble finding it.
*(B) Lack of light made it impossible.
(C) It was too small for a group to fit into.
(D) Items stored by others took up most of the space.

(45) What have archaeologists found in the caves?

Students read and then choose:

(A) Prehistoric desert people.
(B) Migratory animals.
*(C) Food supplies and tools.
(D) Growing plants.

(46) Why does the speaker show a photo to the class?

Students read and then choose:

*(A) To illustrate the size of some objects.
(B) To introduce the next assignment.
(C) To show some artifacts on display at the campus museum.
(D) To demonstrate his photographic ability.

Note: * the keyed option

Each of the three items contains at least one absurd option which, if identified as incorrect by a student using partial knowledge, should be eliminated. Option A for items 44 and 45, and Option D for item 46 were found to be absurd by the two judges and confirmed by item analysis.

Item 44 in Listening Comprehension

Nine (39.1%) test-wise students relied on the ID1 strategy to rule out options as incorrect and then chose from among the remaining option(s). Of these 9 students, 8 reasoned that the food and supplies the prehistoric people carried must have required a lot of space and made the cave dark. Based on this reasoning, they eliminated Options A and C first and then chose either B (6) or D (2), whichever they believed was correct or more logical. The ninth student, although employing the ID1 strategy, approached the item in a different way. She said:

> *I chose B. It was purely guessing because I didn't understand this part. C is dead wrong because I heard that the cave was huge. I'm not sure about A, "They had trouble..." The word "trouble" bothers me. I did not hear this mentioned. It was just my intuitive feeling. So, I randomly guessed between B and D.*

The remaining fourteen (60.9%) students in the test-wise subgroup first caught the key phrase "dark and dusky" from the short talk and then directly connected this key phrase with "lack of light" in Option B (IIB4).

Five (29.4%) of the 17 test-naïve students employed the information that "a lot of food was found in the cave" to rule out Option A and/or other option(s) using the ID1 strategy and then chose B (2), C (1), or D (2). Six (35.3%) students directly chose B based on the key word "dark" they heard from the talk (IIB4). Another two (11.8%)

selected C, claiming they heard the key word "small" mentioned (IIB4). The two remaining (11.8%) test-naïve students randomly guessed.

Item 45 in Listening Comprehension

In contrast to item 44, more students in both the test-wise and test-naïve subgroups responded to item 45 correctly (see Table 11). Further, there was less application of the ID1 strategy. Of the 23 test-wise students, only two (8.7%) students applied the ID1 strategy to eliminate Options A, B, and D as absurd. The remaining 21 (91.3%) students picked C directly, citing that they heard the word "tools" mentioned in the talk (IIB4).

Similarly, among 17 test-naïve students, only one (5.9%) student utilized the ID1 strategy to rule out options A, B, and D. Fifteen (88.2%) students simply chose what they believed they heard in the talk and matched it to the option they selected (14 correct and 1 incorrect). Lastly, one (5.9%) student failed to understand the talk and randomly guessed.

Item 46 in Listening Comprehension

Of the 23 test-wise students, six (26.1%) eliminated options not mentioned in the talk and then chose the option (5 correct and 1 incorrect) based on the key words (e.g., "pictures of tools," "how large," and "something displayed") they believed they heard in the talk. An additional 11 (47.8%) students selected A directly because the word "size" in Option A matched the word "size" they heard in the talk (IIB4). The remaining 6 (26.1%) test-wise students said that they knew the answer (all correct) based on their understanding of the talk.

In contrast, four (23.5%) test-naïve students applied the ID1 strategy to rule out three options (3 correct and 1 incorrect). Four (23.5%) students selected their answers (3 correct) based on the key words they believed they heard. These key words included "picture of tools," "big hand," and "something found from the cave and displayed at the museum." Two (11.8%) test-naïve students randomly guessed (none correctly). One (5.9%) answered the question correctly based on her "hunch". The remaining 6 (35.3%) test-naïve students claimed that they knew the answer (all correct) based on what they understood about the talk.

Items 42, 43, 47, and 50 in Reading Comprehension

The four reading items in the TOEFL Interview Form are based on one reading passage about problems after the American Civil War. After reading the passage, students were asked to choose the best answer from among 4 alternatives.

(42) The word "staggering" in line 1 is closest in meaning to

(A) specialized
(B) confusing
(C) various
*(D) overwhelming

(43) The word "devastated" in line 3 is closest in meaning to

(A) developing
*(B) ruined
(C) complicated
(D) fragile

(47) Why does the author mention a popular song in lines 22-23?

*(A) To give an example of Northern attitude towards the South
(B) To illustrate the Northern love of music
(C) To emphasize the cultural differences between the North and the South
(D) To compare the Northern and the Southern presidents

(50) It can be inferred from the passage that President Johnson pardoned the Southern leaders in order to

(A) raise money for the North
(B) repair the physical damage in the South
(C) prevent Northern leaders from punishing more Southerners
*(D) help the nation recover from the war

Note: * the keyed option

Each of the reading items above contains at least one absurd option that, if known to be incorrect by a student using partial knowledge, should be eliminated. Option A in item 42, Option C in item 43, Option B in item 47, and Option A in item 50 were identified as absurd by the two judges and verified by item analysis based on the entire sample (n = 390).

Item 42 in Reading Comprehension

Item 42 was supposed to measure student's ability to understand the extended meaning of the word "staggering" in the context provided. In addition, knowledge about the literary meaning for the correct answer (D) as well as for the other 3 foils was required to correctly answer the item. Yet prior knowledge about the literary meaning of "staggering" may enhance but not necessarily lead to the correct answer.

Twelve (52.2%) test-wise students eliminated options as incorrect and then chose the answer they believed to be correct. Four of these 12 students successfully ruled out all three distracters; 4 eliminated A and C and chose between B and D (2 correct); and 4 omitted A and B and then selected between C and D (2 correct). For example, one test-

wise student who did not know the meaning of "staggering" explained how she applied the ID1 strategy to arrive at the correct answer:

> *I don't know this word [staggering]. I was thinking that he is talking about the problem they were confronted with. He said that there were more than million people.... Anyway it was very difficult then. I guessed that "confusing" must be incorrect. They knew what should be done. "Various"...this...I don't know. I don't think that "specialized" fits here, either. So, I chose D.*

Two (8.7%) students picked C, thinking that the word "various" fit the context because "there must have been many different things to do after the war." The remaining nine (39.1%) test-wise students selected D directly, claiming that they knew the answer.

One (5.9%) test-naïve student chose D by successfully ruling out the 3 distracters. Eleven (64.7%) students, while able to figure out that the word "staggering" in the context meant "so many and very difficult," appeared to lack sufficient knowledge about the literary meaning of the word "overwhelming." Consequently, they not only eliminated A as absurd but also stayed away from D, the meaning of which was unknown to them, and chose between B and C. For example, one of these students figured out his answer as follows.

> *I am not sure about the meaning of "staggering". It seems to me that it means a trigger or something like that. Anyway, I guessed that there must have been so many tasks for people to do after the war. Therefore, only these two options [B and C] are considerable. Something very urgent anyway. Both B and C mean " so many and disordered". Considering the disordered situations after the war, I selected B.*

Four (23.5%) test-naïve students said that they knew the answer (all correct). One (5.9%) did not know the meaning of the word "staggering" and randomly picked A.

Item 43 in Reading Comprehension

Two (8.7%) test-wise students used the ID1 strategy to eliminate three options, one correctly and the other incorrectly. One (4.3%) student indicated that A and B were opposite options (ID3), and then chose B, arguing that B fit better. The remaining 20 (71.4%) test-wise students reported that they chose B because they knew the answer.

A similar pattern was found with the test-naïve subgroup. Two (11.8%) test-naïve students correctly ruled out 3 distracters (ID1). Two (11.8%) students picked D, based on their understanding of the extended meaning of "devastated", and one (5.9%) selected A because she believed "the word 'developing' is always used to modify the word 'economy'." The remaining 12 (70.6%) test-naïve students chose B, saying that they knew the answer.

Item 47 in Reading Comprehension

Of the 23 test-wise students, 21 (91.3%) used the ID1 strategy to successfully rule out the three options known to be incorrect. For example, one test-wise described how he approached the question:

> *Here I used the deductive reasoning approach. This article has nothing to do with music. Music is not its main topic. Therefore, B can be eliminated. It is unlikely that he would talk about the Northern love of the music. It is irrelevant. It is unlikely, either, that he would compare the cultural differences between the North and the South. As for comparing these two presidents, I think that it has nothing to compare with, since he just mentioned one president. Therefore, I chose A.*

One (4.3%) student eliminated Option B as absurd and chose C because she believed that C covered the meanings associated with A and D (ID4). The last student (4.3%) said that she knew the answer and did not need to employ any test-wise strategy.

Seven (41.2%) of the 17 test-naïve students eliminated options as incorrect and chose the remaining from option(s): four successfully ruled out the 3 foils; two crossed out B and D as wrong and then chose A, which they both believed was better than C; and one first eliminated B and C and then picked her answer (correct) randomly between A and D. The remaining ten (58.8%) students selected their answer (9 correct) based on their understanding of the reading passage.

Item 50 in Reading Comprehension

All but one (95.7%) test-wise student applied the ID1 strategy and successfully eliminated A, B and C as incorrect. Only one student selected Option D directly, claiming that she knew the answer.

In contrast, 8 (47.1%) test-naïve students reported that they used the ID1 strategy to rule out the three foils successfully. The remaining 9 (52.9%) students directly selected their answer (8 correct) based on what they understood about the reading passage.

To summarize the differences between the test-wise and test-naïve students when responding to the TOEFL items susceptible to ID1, the test-wise students tended to be more cautious about their final selection of an answer, and tended to use the ID1 strategy more frequently and, more importantly, more effectively than their test-naïve counterparts. In many cases, it was the quality rather than the quantity of use of the test-wise strategies that accounted for the differences in performance between the two subgroups. Item 42 in Reading Comprehension is an example. While the two subgroups were close in use of other test-wiseness strategies for item 42, the proportion (70.6%) of the test-naïve students who employed the ID1 strategy exceeded the proportion (52.2%)

of the test-wise students who did likewise. Despite this, only one out of 12 test-naïve students who used the ID1 strategy managed to successfully eliminate 3 distracters and obtain the correct answer. The remaining 11 students all ruled out the options including the correct answer. In contrast, 8 out of the 12 test-wise students who applied the ID1 strategy successfully eliminated 2 or 3 distracters and correctly answered the question. As the result, the *p*-value (73.9%) of the correct answer for test-wise students was substantially higher than that (29.4%) of their test-naïve counterparts.

The test-wise student's application of the ID1 strategy often served double purposes: (1) to rule out options as incorrect and choose from the remaining option(s) when they were not so positive about the correct answer on the basis of their knowledge or understanding alone; and (2) to confirm their choice as correct even when they felt that they knew the answer.

In contrast, test-naïve students tended to rely more on their knowledge, particularly their prior knowledge and "common sense," or on what they believed they understood about the talk/passage provided. When facing an easy item, the test-wise and test-naïve students were equally or nearly equally able to answer the question correctly. However, when confronted with a difficult item to which they did not know the answer, the test-naïve students tended either to give up easily and randomly guess their answers or to choose an answer using their knowledge alone, no matter whether this knowledge was sufficient or not.

Items Susceptible to ID2: Similar Options

One item, item 44 in Reading Comprehension, was identified as susceptible to the ID2 cue. That is, Options B and D appear to be the same or similar in meaning and, therefore, nether can be the correct answer.

Item 44 in Reading Comprehension

As mentioned previously, item 44 is based on the passage about the problems after the American Civil War. Following the presentation of the reading passage, the students were asked to choose the best answer from among 4 options.

(44) According to the passage, which of the following statements about the damage in the South is correct?

*(A) It was worse than in the North.
(B) The cost was less than expected.
(C) It was centered in the border states.
(D) It was remedied rather quickly.

Note: * the keyed option

Item 44 in Reading Comprehension had been identified as susceptible to the ID2 cue first by judgment and then by the empirical evidence obtained from the item analysis based on the sample students (see p. 110). Nevertheless, the results of "think-aloud" protocol analysis revealed that the solution strategies adopted by both test-wise and test-naïve subgroups were not as expected. Of the 23 test-wise students, four (17.4%) picked A citing "common sense." Two (8.7%) students selected B on the basis of the word "less spectacularly" in the passage (IIB4). Two (8.7%) students applied the ID1 strategy: one chose A following the elimination of B, C, and D as incorrect; and the other first ruled out B and D and then chose C randomly between the remaining Options A and C. The

remaining 15 (65.2%) students directly selected the answer (14 correct) based on their understanding of the passage.

A similar pattern was observed among the 17 test-naïve students: four (23.5%) picked the answer (1 correct) citing "common sense." One (5.9%) student used the ID1 strategy to rule out A and C and then chose randomly between B and D. One (5.9%) student randomly guessed (wrong). The remaining 11 (64.7%) test-wise students selected A based on their understanding of the passage. In short, no student was found to recognize and eliminate the two similar options (B and D).

One of major differences between these two subgroups seems to lie in the quality of the "common sense" each subgroup cited. Test-wise students' "common sense" appears to make more sense than test-naïve students'. For example, one test-wise student cited her common sense: "Since the war was fought in the South, the damage must be worse there than in the North." Her "common sense" was more or less the same as the "common sense" the other 3 test-wise students cited. In contrast, one test-naïve student cited her common sense: "The American Civil War must have occurred along the border between the North and the South. Naturally, the states along the border suffered more." Two additional test-naïve students cited this similar "common sense." It appears that the differences between the two subgroups in common sense were attributed to the differences in knowledge each subgroup possessed. The students classified as test-wise seemed to be more academically knowledgeable than the students classified as test-naïve. This finding is consistent with what Rogers and Bateson (1994b) reported.

Items Susceptible to ID3: Opposite Options

There were four listening items and one reading item that contained a pair of opposite options. The results for the four listening items are presented and discussed first, followed by the results and discussion for the reading item.

Item 27 in Listening Comprehension

Item 27 is based on a brief conversation between a man and a woman.

Students hear:

(Woman): Have you heard anything about the new professor?
(Man): Just that she's no pushover.
(Narrator): What does the man say about the professor?

Students read and then choose:

(A) She works very hard.
*(B) She is very strict.
(C) Her classes fill up quickly.
(D) It's easy to get good grade in her courses.

Items 34, 35 and 37 in Listening Comprehension

Items 34 through 37 are based on a conversation between a man and a woman.

Students hear:

(Narrator): Questions 34 through 37. Listen to a conversation between two friends.

(Man): Hey, how was your trip?
(Woman): Wonderful. I spent most of my time at the art museum. I specially liked the new wing. I was amazed to hear the guide explain all the problems they had building it.
(Man): Right. I just read an article that went on and on about the cost – ninety million total, I think.
(Woman): Yeah, the guide mentioned that. You could see they spared no expense.
(Man): Hmm. It looked really unusual, at least from what I saw in the picture.
(Woman): It is. The basic design is two triangles. In fact, there are triangles all over – the paving stones in the courtyard, the skylights, and even a lot of

the sculptures. One sculpture is a mobile. It's in the courtyard and it's made of pieces of aluminum that move slowly in air. It's really impressive.

(Man): That was in the article, too. It said the original was steel, and it weighed so much that it wasn't safe to hang.

(Woman): Right. They did it over in aluminum so it wouldn't come crashing down on someone's head.

(Man): You know, the article went into that in detail. There was even an interview with the sculptor.

(Woman): I'd like to read that. Would you mind if I borrowed the magazine sometime?

(Man): No, I wouldn't mind, if I haven't thrown it out yet.

(34) What did the woman think of the new wing of the museum?

Students read and then choose:

*(A) She was impressed by it.
(B) It was a waste of money.
(C) She was amazed it had opened so soon.
(E) She didn't like it as much as the other wings.

(35) How had the man learned about the museum?

Students read and then choose:

(A) He took a tour of the city.
*(B) He read about it.
(C) He wrote an article about it.
(D) He worked there as a guide.

(37) What was the problem with the original mobile?

Students read and then choose:

*(A) It was made of aluminum.
(B) It wasn't large enough.
(C) It wouldn't move in the wind.
(D) It was too heavy to put up.

Note: * the keyed option

Item 27 in Listening Comprehension

In item 27, Options B and D were identified as opposites, one of which was the correct answer. Of the 23 test-wise students, only one (4.3%) knew the word "pushover" and picked B directly. Seven (30.4%) students identified B and D as opposites and selected between these two options (6 correctly) in the following ways: 4 chose the answer based on speaker's tone, two guessed randomly, and one cited her "common sense" that "a teacher is usually strict." Eight (34.8%) students utilized the narrator's question as a clue and ruled out option(s) as irrelevant and then selected the option which they felt was more like students' typical comments about the professor. Of these 8 students, two eliminated C and D as irrelevant and chose between A and B (1 correctly); 4 also eliminated C and D as irrelevant and then eliminated A (ID1) because they thought "whether a professor worked hard or not was irrelevant to students;" and two picked B, citing their experience that professors were usually strict with their students. One (4.3%) student guessed B according to her "intuition" and four (17.4%) students picked the answer (1 correct) by establishing some connection between "pushover" and one of the 4 alternatives (IIB4). For example, one test-wise student said:

> *I didn't get it. I heard the phrase "push...pushover." I am not familiar with this phrase. I thought that professor must be very strict if she pushed over her students all the time. I chose B.*

For another test-wise student who used the same approach, the word "pushover" meant something else:

> *He said that the class was packed with students.... I got the first sentence but failed to catch the second one. I just caught the word, something like "pushover" and "seats." I thought that, if people pushed each other, the class must be full. So, I decided to select C.*

The remaining two (8.7%) students chose B on the basis of the second speaker's tone alone. It is noteworthy that, although students did not know the literal meaning of the word "pushover," the tone of the speaker provided a strong clue for them. For example, one test-wise student who identified B and D as opposites explained:

> *First of all, these two [B and D] are somewhat opposite: one is strict and the other is very easy-going. So, I bet on these two. TOEFL teacher told us to choose something positive when guessing. But I felt it should be "strict", because, based on the man's tone, I felt so. You can tell how the second person thought about somebody or something by the tone. This is one of the characteristics of the TOEFL.*

Another test-wise student explicitly articulated how she used the speaker's tone:

> *From listening, I just felt that this sentence might mean something bad. From the way the lady asked the question, and based on his tone, I felt that it must involve a sort of unfavorable comment. So I guessed D was not likely. It should be something opposite D. Rather than it is easy to pass her course, it should be more difficult. Therefore, I thought to choose "This professor is strict". I chose B...because they only talked about the professor. C is just about her classes. I thought what this man said had nothing to do with classes. Therefore, I eliminated C. As for A, judging from this male student's tone, it [no pushover] sounds like something bad. If it was A, then it meant to speak highly of this professor. So, I chose B. She is strict and it is difficult to pass her test.*

Of the 17 test-naïve students, two (11.8%) identified B and D as opposites and then chose B using their common sense. Two (11.8%) students first ruled out A as irrelevant, and C as grammatically incorrect or D as absurd, and then guessed between the remaining two options (both correct). Eight (47.1%) test-naïve students relied upon their common sense to select the answer (6 correct). Six of these eight students cited their common experience that "a professor is often strict and harsh." Three (17.6%) students chose the answer (2 correct) based on the second speaker's tone alone. Of the remaining two (11.8%) students, one randomly guessed C, and the other picked B according to her "intuitive feeling."

Item 34 in Listening Comprehension

In item 34, Options A and D were semantically opposite. Nevertheless, both judges felt that the correct answer was more likely between Options A and C because the two options, while containing a synonymous phrase (i.e., "she was impressed/amazed..."), were different in meaning. In this case, based on the judges' experience with TOEFL listening items, the correct answer was likely to be a paraphrase of what was said in a conversation whereas the distracter(s) often contained a word/phrase that was spoken by a speaker. Their judgment and reasoning was later confirmed by the empirical evidence. The results of item analysis of the TOEFL based on the full sample revealed that 344 (88.2%) and 29 (7.4%) of the 390 Chinese students selected A and C, respectively, and their corresponding TTW mean scores were 13.4 and 11.7, both higher than the TTW mean scores for either of the remaining two groups who chose B or D (see Appendix E). Further, the empirical evidence was also found from the protocol analysis of the students' think-aloud data.

Of the 23 test-wise students, 3 (13.0%) recognized A and C as opposites and then selected A claiming that they heard the word "impressed" or something similar from the conversation (IIB4). One of them explained how he figured out the answer:

> *Prior to listening to the tape, I quickly scanned the options. I guessed that the question might be about how she thought about something. If so, I knew that the answer must be between these two options [A and C] because they both contain a similar phrase "she was impressed/amazed" but differ from each other in meaning. According to my experience, in most cases, the answer is likely hidden between this kind of options. While listening to the conversation, I was looking for the relevant information. As soon as I heard her saying "Wonderful... I was impressed" or something like that, I knew the answer must be A.*

Two (8.7%) students selected the answer (1 correct) based on the words they caught from the conversation: one claimed to hear the word "impressed"and the other said that he heard the word "amazed" mentioned. Three (13.0%) students chose Option A based on the woman's tone. As one test-wise student said: "I chose A because she sounds excited, though I didn't catch the exact word she used." The remaining 15 (65.2%) students chose their answers (all correct) based on what they understood about the conversation.

No test-naïve students recognized A and C as opposites. Three (17.6%) test-naïve students responded to the question based on the key word they caught from the conversation (IIB4): two selected C because they heard the word "amazed," and one picked A, claiming that he heard the word "impressed." The remaining 14 (82.4%) students said that they knew the answer and selected their answer (all correct) accordingly.

Item 35 in Listening Comprehension

In item 35, Options B and C were identified as opposites, one of which was the correct answer. Three (13.0%) test-wise students first recognized B and C as opposites and then decided to choose B because they believed that to read an article was more likely to occur in daily life than to write an article. For example, one of these students said:

> *I did not quite understand the conversation. So, I have to use deductive reasoning. I knew the answer is more likely between B and C. The difference between the two options lies in whether he read or wrote an article. Between B and C, I think that B is more likely to happen in everyday. Therefore, I selected B.*

Another 5 (21.7%) test-wise students relied on the key word "article" or "magazine" they caught to eliminate A and D as irrelevant (ID1), and then ruled out C as less likely in everyday life (ID1). One student explained:

> *I'm not so sure about item 35. All I got was the key word "article." So, I ruled out A and D. Between B and C, I chose B. I guessed the probability of getting it right is high because it is more likely for ordinary people to read rather than to write an article.*

A similar approach was used by another test-wise student:

> *At first he said something like "saw a picture about it." I did not catch the rest of the conversation but I knew that the lady wanted to borrow the magazine. This [A] only fits the lady. This [C] is not correct because he wouldn't have such an expertise. Also that [D] is not a fact. I used deductive reasoning to eliminate the impossible answers.*

One (4.3%) student randomly guessed, since she missed a big chunk of the conversation. The remaining 14 (60.9%) test-wise students selected B, saying that they knew the answer.

The pattern for test-naïve students' solution strategies used in responding to item 35 was slightly different from that for test-wise students. No one recognized the pair of opposites. Three (17.6%) test-naïve students employed the key word "article" or "magazine" they heard to eliminate Options A and D as incorrect, and then chose the answer (3 correct) from the remaining options based on their common sense. In this case, the 3 test-naïve students applied an approach similar to what the test-wise students used. For example, one test-naïve student described how she arrived at the correct answer:

> *I didn't fully understand it. I got the answer indirectly. I reasoned it out. I heard that he lent the magazine to her. Then I thought about C. Yet not every American likes to brag. I left C undecided. A is not right for sure, and neither is D. B is most likely. It is more common that a person read about it than wrote about it. To write is less possible than to read. Therefore, I chose one that is more common and inclusive. I'm 80% sure about my decision.*

The remaining 14 (82.4%) students directly chose their answers (11 correct) based on what they understood about the conversation.

For both test-wise and test-naïve students, the use of the key word "article" or "magazine" seemed to play an important role in their "educated guessing." When not so sure about the correct answer, they tended to relate the key word "article" or "magazine" they heard from the conversation with the word "read" in B and/or the word "write" in C. This finding may provide an explanation for the test behavior of the full sample (n = 390) of the students on item 35 that 85.1% of the 390 students picked either B or C and avoided A and D.

Item 37 in Listening Comprehension

Item 37 is based on the same conversation as items 34 and 35. It also contains a pair of opposites, Options A and D. Of the 23 test-wise students, none claimed that she/he fully understood the relevant part of the conversation. Nonetheless, only one (4.3%) student gave up and randomly guessed. The remaining 22 (95.7%) students made a full use of the key word(s) they caught from the conversation and reasoned out their answers (19 correct). Of these students, 12 grasped the scattered but relevant information such as "it was made of steel before and aluminum now", identified A and D as opposites, and then either ruled out A as incorrect (6 students), or related the word "problem" in the question with the word "steel" in the conversation and the phrase "too heavy" in Option D (6 students). Following are two examples to illustrate how test-wise students, while unable to fully understand the conversation, managed to figure out the

correct answer. A test-wise student described how he made a use of the scattered information he heard and selected the answer:

> *The question is about the problem. I didn't get the whole conversation clearly. So, I had to guess the answer. I felt C was not likely because if an automobile was unable to move in the wind, then it can't be called an automobile. Then "it wasn't large enough..." As for this, I didn't hear any relevant information. So, the choice must be between A and D. I did hear they mention the word "aluminum". A thing made of aluminum is light... so, the problem must be its opposite —— too heavy. Therefore, I chose D.*

Another test-wise student explained how he capitalized on relevant information obtained from item 36, ruled out B and then chose between A and D:

> *They asked something about the original one. I guessed that it must be heavy. Well, since I chose C in the previous item [item 36], i.e., "they are similar in shape," I would not consider any option involving size change such as B. I just followed the same thought stream. In addition, I felt that A is opposite to D because anything made of aluminum must be light. The old one usually was not as good as the new one and it must have some weakness. The aluminum one is always lighter and better. As for C, it is impossible that it wouldn't move in the wind. Regardless of whether it is an old or new model, it wouldn't have any value if it could not move. The differences between them [old and new models] might be that one moves faster than the other.*

Two students eliminated A and B as wrong using the key words "steel," and then chose D claiming "since it was steel, it must be heavy." Another five students directly chose their answers by connecting the key word "steel" in the conversation with "it was too heavy" in Option D (4 students) or with "it wouldn't move in the wind" in Option C (1 student). One student ruled out A and C citing his common sense and then guessed D. The remaining two students, while both making a use of the key word "aluminum", used it differently. One chose A because both the conversation and Option A contain the word "aluminum". The other ruled out A as incorrect and then guessed D from among the remaining 3 options.

In contrast, 3 (17.6%) of the 17 test-naïve students chose their answers (all correct) based on their understanding of the conversation. Three others (17.6%) failed to catch the relevant information and randomly guessed (all incorrectly). The remaining 11 (64.7%) test-naïve students made a use of the key word(s) they heard: 3 first recognized A and D as opposites and then chose D by relating "too heavy" in the option to the word "steel" they heard; 3 directly chose the answer (all incorrect) containing the key word they believed they heard from the conversation; one heard the word "aluminum" but deliberately picked its opposite----D; and 4 first ruled out C as absurd and then used the obtained information " it is made of aluminum now" to eliminate A and select D.

Item 49 in Reading Comprehension

Item 49 is based on a reading passage pertinent to problems after the American Civil War. After reading the passage, the students were asked to answer the question:

(49) Which of the following can be inferred from the phrase "...it was unlikely that a jury from Virginia, a Southern Confederate state, would convict them" (lines 25 – 26)?

(A) Virginians felt betrayed by Jefferson Davis.
(B) A popular song insulted Virginia.
*(C) Virginians were loyal to their leaders.
(D) All of the Virginia Military leaders had been put in chains.

Note: * the keyed option

Options A and C were identified as opposites, one of which was the correct answer.

All of the 23 test-wise students selected either Option A or C and avoided B and D. Two (8.7%) students recognized A and C as opposites and then selected C based on what they understood about the text/stem. One of these two students described how he applied deductive reasoning to determine his answer:

> *I am not so sure. I used deductive reasoning approach again. Virginia is one state in the South. They can't do something harmful to the Southern leaders. I don't know the word "convict," but I guessed that it means punish. B is not right because it is too specific to be the answer. A and C are opposite. I believe that if the two options are opposite, then one of them must be the answer. According to my experience, as well as what teachers taught, if the two options are opposite, then the correct answer may be within one of them. In this case, A and C are opposite. And later they released them [the Southern leaders]. Therefore, they were loyal to their leaders. I chose C.*

Six (26.1%) students eliminated B and D as absurd and then ruled out the third option (5 correctly) based on their understanding of the stem. Two (8.7%) students said that they knew the answer (both correct). The remaining 13 (56.5%) students all chose Option C claiming that there was a logical connection between "unlikely...convict them" in the stem and "loyal to their leaders" in Option C. As one of these test-wise students articulated:

> *Virginians were Southerners and he [Jefferson Davis] was also a Southerner. Therefore, C is more likely. According to the text, they were released at end. The jury may play a key role in their release. According to the geographical location, they were all Southerners. Naturally, they were loyal to their former leaders.*

In contrast, two (11.8%) test-naïve students selected D: one randomly guessed, and the other misunderstood the stem and thought that she was being asked "which of the options was unlikely." Ten (58.8%) students claimed that they knew the answer and directly chose A (8 students) or C (2 students). Two (11.8%) test-naïve students first eliminated two options and then chose C either citing "common sense" or randomly. For example, one of these students described how she struggled to make a choice between C and D:

> *I am not so sure about the answer. I eliminated B first because I thought it was absurd. I didn't consider A, either. I was focussed on C and D. I chose C first. Then I read D, but I could not fully get it. I am not sure.*

So, I chose C. No, this time, I give up. However, I have to guess from C and D, do I? I'm choosing C. C is better, I think.

The remaining 3 (17.6%) test-naïve students selected Option C because they thought that there was a logical connection between Option C and the stem (IIB4).

To summarize the differences in test behaviors between the test-wise and test-naïve subgroups when responding to the TOEFL items susceptible to ID3, the test-wise subgroup appeared to be more capable of recognizing and making use of the ID3 cues than their test-naïve peers. When dealing with the listening items of the TOEFL, the test-wise subgroup were more able to eliminate distracters and focus their attention on the remaining options for a possible correct answer than did the test-naïve subgroup.

However, recognizing and choosing between the two opposites was not the most used test-wiseness strategy for the Chinese students, particularly for the listening items. Rather, finding a connection between an option and the key word heard from the conversation/talk (IIB4) or a combination of IIB4 and ID1 was more commonly applied. In terms of using IIB4 or combination of IIB4 and ID1, the test-wise students again outperformed the test-naïve students.

Items Susceptible to ID4: Options that Converge

There was one item, item 41 in Reading Comprehension, for which the ID4 test-wiseness strategy could be used.

Item 41 in Reading Comprehension

(42) What does the passage mainly discussed?

Students read and then choose:

(A) Wartime expenditures.
*(B) Problems facing the United States after the war
(C) Methods of repairing the damage caused by the war
(D) The results of government efforts to revive the economy

Note: * the keyed option

Item 41 is based on the same reading passage about problems after the American Civil War. The correct option (Option B) encompasses the meanings of Options C and D. Four (17.4%) test-wise students selected B because they felt that B covered all the other options except A. One of these students explained:

> *Option B best summarizes the text, I think. It also covers the meaning of the other options. So, I chose B.*

Eleven (47.8%) test-wise students applied the ID1 strategy and successfully eliminated the three distracters. For example, one test-wise said:

> *The first paragraph is about the problems they were facing rather than how they solved these problems. So, I eliminated D first. In addition, [the problems] were not just in economy. C is not right because they mainly talked about the problems rather than repairing the damage. A is not correct in the first part of the statement.*

Nine of the 11 students in this group combined ID1 with ID4 together to figure out or to confirm their answers. For example, one student commented:

> *I eliminated A and D first. A and D are not mentioned at all. They are just a part of the whole. C is also a part of it. The word "problems" covers all of things talked about in the text and in other 3 options. So, B is the best answer.*

The remaining 8 (34.8%) test-wise students all reported that they picked the answer (7 correct) based on their understanding of the text.

Three (17.6%) of the 17 test-naïve students employed the ID1 strategy and successfully ruled out 3 foils. The remaining 14 students selected their answers directly, claiming that they knew the answer (11 correct).

The Prediction Strategy

As Table 12 shows, one prevailing strategy applied by the interview subsample in responding to the TOEFL listening items was to predict questions using the 4 options provided before listening to the tape. All of the students in the interviews admitted that this strategy was one of major test-taking skills they were trained to learn in class. They cited that prediction could help them roughly know beforehand what topic/theme of the conversation/talk and up-coming question might be so that they could listen selectively for relevant information and key words. During the interviews, some students were even concerned whether reading options ahead is allowed in the real TOEFL testing situation. One test-wise student explained why it was so important for her to predict:

> *I can't do well without knowing beforehand what is main theme of the conversation. Reading options ahead, I would know what to listening to, which option is the potential answer, and which one I should ignore. At least, reading ahead and prediction could release my test-anxiety and help me get ready mentally and psychologically for the up-coming conversation.*

Differences between the test-wise and test-naïve subgroups were found in using the "prediction" strategy to deal with listening items. In general, 82.6% to 100% of the 23 test-wise students applied or claimed to apply the "prediction" strategy when

responding to each of the listening items listed in Table 12 whereas 67.4% to 88.2% of the 17 test-naïve students reported that they did the same. The percentage of the test-wise students reported that they used the "prediction" strategy was consistently higher than the percentage of the test-naïve students who claimed that they utilized the same strategy.

Further, differences between the two subgroups were observed in the ability to use the "prediction" strategy. For example, one test-wise student elaborated how she used the "prediction" strategy to answer item 36 in Listening Comprehension:

> *I usually guessed the question and then pre-responded before I heard the question. I read options before and after listening. I tried to match options with what I heard from the conversation and eliminated those I did not hear. For example, in item 36, according to the options, I figured the question might be about the similarity among something. Yet I did not hear them mention A, B and D. I knew it must be C. I used the interval between questions and tried to do next one before the question [came up], if possible. Then I looked for those that looked like an answer, and compared them with what I heard, and then I would know the answer.*

Another test-wise student responded to item 46 in Listening Comprehension using a similar approach:

> *I first read the 4 options. I predicted the question and decided it [the correct answer] was A. Later, I heard that he [the speaker] said, "you can feel the size of that object by hand." That is to say, according to the content, it should be A.*

When the student's predictions were incorrect, they also knew how to adjust and make corrections accordingly. For example, when responding to item 37 in Listening Comprehension, one test-wise student first circled Option A before the question was asked, and then changed to D after she heard the question. When asked why she did so, she explained:

> *I thought that it was made of aluminum. They mentioned aluminum. The remaining options are about reasons. I thought that, since they mentioned aluminum, they might ask what the car, or automobile, is made of. If so, it*

should be A. But the question they asked was what the original problem was. So, I have to change my answer.

In contrast, the test-naïve students, although aware of importance of reading ahead and prediction, were somehow unable to fully implement this skill. For example, one test-naïve student, after describing her approach to the last listening item, hesitated for a second, then said:

Be frank with you, I did not use much test-taking skills. I did not read ahead and predict. It may distract my attention and focus.

This student was not the only one who felt the same way among the test-naïve students. Another two test-naïve students also claimed that they never or seldom predicted or read ahead options before listening to the TOEFL listening items.

Non-susceptible Items

Eight TOEFL items in the Interview Form were considered to be non-susceptible to test-wiseness. These items can be classified into three categories: (1) items identified by the two judges as susceptible to a certain test-wiseness element but not supported by empirical evidence obtained from the item analysis; (2) items identified by the two judges as non-testwise susceptible but found vulnerable to the test-wiseness strategies Chinese students commonly applied in responding to the TOEFL items in the interview; and (3) items found to be clean and not susceptible to test-wiseness by both the subjective judgment and empirical evidence. The following presentation is focussed on how the test-wise and test-naïve students performed when responding to the items identified as non-susceptible to test-wiseness. The discussions are presented by category.

First Category

Items 10 and 11 in the Listening Comprehension subtest belonged to the first category. Item 10 was identified as susceptible to the opposite-option cues (ID3) and item 11 as susceptible to several test-wiseness elements including the absurd option (ID1), and longer and more qualified option (IIB1a and IIB1b). However, these judgments were not supported by the empirical evidence obtained from the item analysis of the TOEFL based on the full sample (see Appendix E).

Item 10 in Listening Comprehension

Students hear:

(Man): Are you going home for winter vacation?
(Woman): I've agreed to stay on here as a research assistant.
(Narrator): What can be inferred about the woman?

Students read and then choose:

(A) She'll be travelling during winter break.
*(B) She'll be working during vacation.
(C) She's looking forward to going home.
(D) She wants to hire another research assistant.

Item 11 in Listening Comprehension

Students hear:

(Man): Hello.
(Woman): Hello. This is Dr. Gray's office. We'er calling to remind you of your 4:15 appointment for your annual checkup tomorrow.
(Man): Oh! Thanks. It's a good thing you called. I thought it was 4:15 today.
(Narrator): What does the man mean?

Students read and then choose:

(A) He's glad he called the doctor.
(B) He wants to change the appointment.
(C) He can't come until 4:15.
*(D) He was confused about the date of the appointment.

Note: * the keyed option

Item 10 in Listening Comprehension

For item 10, Options A and B, while almost the same in the sentence structures and vocabulary used, completely differed in meaning due to the different main verbs used for each sentence. The two judges felt that, in item 10, this pair of the options that looked alike syntactically but differed semantically served as an "opposite-option" cue which implied that one of the options was likely to be the correct answer. Nonetheless, the results of the item analysis revealed that 331 (84.9%) of the 390 Chinese students selected B whereas only 6 (1.5%) students chose A and their TTW mean score was low ($\bar{X}_{TTW} = 10.8$) (see Appendix E). Unexpectedly, the second most popular option was D, which attracted 35 (9.0%) students.

In the follow-up interviews, all of the 23 test-wise students reported that they read the 4 options in advance and predicted the up-coming question. Twenty-one (91.3%) test-wise students directly picked B, claiming that they understood the conversation and knew that B was, therefore, the answer. The remaining two (8.7%) test-wise students first predicted the question and then applied a unique test-taking strategy which was not listed in Millman et al.'s (1965) Taxonomy of Test-wiseness Principles. This strategy, like a reversal of IIB4, can be best described by a test-wise student:

> *I knew that the coming question would be something like "What she will do." I focussed my attention on the woman. I heard her saying something about ... research assistant. I was not quite sure. Option D contained "research*

assistant" but I guessed it was not the answer. It was too easy to be true. My teacher told us, "unless 100% sure, never pick an option containing one or two words/phrases or a homonym/similar sounding word spoken in the conversation. They were more likely to be a trap. Instead, choose an option that paraphrased what was heard." So, I checked the 4 options. It was not A or C. I heard the phrase "going home in winter." So, the only option left was B. I think that B was the answer.

In contrast, 13 (76.5%) of the 17 test-naïve students read the options ahead and tried to predict the question prior to listening. Fourteen (82.4%) test-naïve students said that they understood the conversation and knew that B was, therefore, the correct answer. One (5.9%) student chose C randomly. The remaining two (11.8%) test-naïve students picked D, claiming they heard the phrase "research assistant" mentioned in the conversation (IIB4). Their approach may shed some light on explaining why D was the second most popular option among the Chinese students.

Item 11 in Listening Comprehension

Unlike Options B, C, and D, which were all associated with time and appointments, Option A appeared to be semantically unique. In this case, Option A was considered the least likely to be the correct answer. Instead, Option D looked more like the correct answer because it was longer (IIB1a) and represented a higher degree of generalization (IIB1b) than the other 3 options. Nonetheless, the results of item analysis showed that the item met all major psychometric requirements—the *p*-values approximately evenly distributed among the 3 distracters, and a moderately high point-biserial coefficient (r_{pbis} = .49) for the correct option (see Appendix E).

All of the 23 test-wise students read ahead the 4 options and predicted the content of the up-coming conversation. One test-wise student described:

> *Before listening, I read the 4 options and guessed that the content of the conversation might be about the appointment ...perhaps with...with a doctor. So, I focussed my attention on the specific time like 4:15. Later, I heard the lady saying "4:15 tomorrow." I thought that they might ask what time was the appointment. But they didn't. They asked, "What does the man mean?" The man said that he...he thought it was 4:15 ...today. He forgot the date. So, the answer must be D, "He was confused about the date."*

Nineteen (82.6%) test-wise students claimed that they knew the answer (all correct). Three (13.0%) students successfully eliminated 3 options as incorrect. One of them reasoned:

> *I don't think that A and C were right. He didn't call the doctor. It was the doctor who called him. I didn't hear C, either. Between B and D, I chose D because it made more sense to me.*

The last test-wise student (4.3%) selected A based on the information she obtained from the conversation (IIB4). She said:

> *I heard the man say something like "Oh! Thanks! Good!" Judging by his tone, I knew that he was happy. Therefore, I chose A.*

Of the 17 test-naïve students, 13 (76.5%) pre-read the options and predicted the topic of the conversation. All claimed that they knew their answers: 14 selected D and 3 picked B based on their understanding of the conversation.

In sum, items 10 and 11, while not as susceptible as expected, were vulnerable to some test-taking strategies such as the "prediction" and the stem-option cue (IIB4). With assistance of the "prediction" strategy, the students were able to predict the content/topic of the up-coming conversation. Doing so enabled the students to do selective listening--concentrate on catching certain key words/information which had been pre-determined as relevant to the correct answer in options provided. Also, the unique listening strategy the

two test-wise students applied when responding to item 10 was worth noticing. The strategy can be summarized as: choose the paraphrase of the speaker's statement(s) rather than an option containing one or two words/phrases spoken in the conversation. Further research needs to be done to confirm the existence of such a strategy.

Second Category

Items 36, 42, and 43 belonged to the second category. They were identified as non-testwise susceptible items. However, these items appeared to be susceptible to certain test-wiseness strategies, such as IIB4 and the "prediction" strategy, which the subsample of the students frequently employed when responding to the TOEFL listening items.

Item 36 in Listening Comprehension

Students hear:

...

(Man): Hmm. It looked really unusual, at least from what I saw in the picture.

(Woman): It is. The basic design is two triangles. In fact, there were triangles all over ---- the paving stones in the courtyard, the skylights, and even a lot of the sculptures. One sculpture is a mobile. It's in the courtyard and it's made of pieces of aluminum that move slowly in the air. It's really impressive.

...

(36) According to the woman, what do the paving stones, skylights, and mobile have in common?

Students read and then choose:

(A) They came from the original wing.
(B) They'er made of the same materials.
*(C) They'er similar in shape.
(D) They were designed by the same person.

Items 42 and 43 in Listening Comprehension

(Narrator): Questions 42 through 46. Listen to part of a talk given in an anthropology class.

(Man): Today's lecture will center on prehistoric people of the Nevada desert. Now, most of these prehistoric desert people moved across the countryside throughout the year. You might think that they were wandering aimlessly – far from it! They actually followed a series of carefully planned moves. Where they moved depended on where food was available – places where plants were ripening or fish were spawning.

...

(42) What is the main subject of this talk?

Students read and then choose:

(A) Rock formation in the Nevada desert.
(B) Graduate studies in anthropology.
(C) Excavation techniques used in archaeology.
*(D) Prehistoric desert people of Nevada.

(43) What point does the speaker make about the prehistoric people of the Nevada desert?

Students read and then choose:

*(A) They planned their migrations.
(B) They didn't travel far from their base camps.
(A) They hid from their enemies in caves.
(B) They planted seeds near their camps.

Note: * the keyed option

None of the items above appeared to be susceptible to test-wiseness elements and the results of the item analysis of the TOEFL did not show any major psychometric problems for those items, either. Nevertheless, some interesting test-taking behaviors

were observed when the subsample of the students responded to these items in the interviews.

Item 36 in Listening Comprehension

Of the 23 test-wise students, 14 (60.9%) selected C mainly based on the word "triangle" they caught from the conversation (IIB4). One test-wise student's approach was typical of the responses of these students:

> *I didn't catch the whole conversation. But I heard them say "triangle...the basic shape of the two things is triangle." I guessed that the shape should be basically similar. So, I chose C.*

Two (8.7%) test-wise students eliminated 3 options (ID1), one of which was the correct answer. One of them explained his approach:

> *I didn't quite understand. But I heard a few words like "sculpture," "skylight," etc. Then, I checked the options. A was not likely because "the original wing" was stolen [steel]. C was not correct for sure. Three objects were not similar in shape, nor likely to be designed by the same person. It was possible that they were made of the same material. I did hear the word "aluminum." Hence, I guessed that B was the right answer.*

One (4.3%) additional student chose B based on some scattered information he obtained from the conversation:

> *I didn't get it. I forgot the relevant information. The reason for me to select B was that I knew the material was changed for a triangle because it was too heavy. I thought that they originally must have been made of the same material. Later, some of the parts were replaces by some other material because they were too heavy or too large. So, I selected B.*

Two (8.7%) test-wise students first used the relevant information in item 37 to eliminate Option B in item 36 (ID5), then relied on the information obtained from the conversation to rule out D (ID1), and finally chose from among the remaining options (both correct). One student explained:

> *B is not likely. If it is made of the same material, it would...I don't know. I could not figure out what the word "wing" here is referred to. I interpreted it as something like a kite. When I read the next set of options [item 37], I realized that B is not correct because the next question is about the differences between the new and the old. It is possible to keep the original shape unchanged and only make changes in the material and in other aspects. D seems to be impossible. To be designed by the same person does not seem to be right. I don't think that they mentioned this person, either. However, between A and C, I can't tell. I chose C based on my gut feeling.*

The remaining 4 (17.4%) test-wise students randomly guessed the answer (all wrong) because they failed to catch any information related to the question.

In contrast, 3 (17.6%) test-naïve students selected C based on the word "triangle" they heard from the conversation (IIB4). One test-naïve student said:

> *Since I didn't have enough time to read the 4 options before listening, I just read first two and paid attention to these two [Options A and B] while listening. I didn't understand most of the conversation. It seemed that they said something like they were made of aluminum. So, I picked B.*

Four (23.5%) additional students utilized the words they heard to establish a direct or indirect connection between the stem and an option (IIB4): two related Option B with the word "aluminum;" one found Option A containing the word "wing" mentioned in the talk; and the last student linked the word "sculptures" she heard with Option B because she believed, "all sculptures were made of the same material--plaster." Another test-naïve student chose D claiming that she heard the woman listing a number of the same kind of things, which she reasoned must be designed by the same person. Four (23.5%) test-naïve students used the information obtained from the conversation to eliminate two options and then chose randomly from the remaining two options (ID1; all incorrect). Of these students, two ruled out B and C, and the other two eliminated B and D, and A and D, respectively. One (5.9%) test-naïve student picked D claiming that she knew the answer. The remaining 4 (23.5%) randomly guessed.

Items 42 in Listening Comprehension

All of the 23 test-wise students read the 4 options before listening and predicted the up-coming question. And all of them reported that they got the answer from the first sentence of the speaker. One of typical responses was:

> *Before the conversation started, I scanned the 4 options. This was the first item and the options all read like a title. I guessed that the question they were going to ask would be something like "what is the topic." Our teacher taught us to pay a special attention to the very first sentence, which usually contained the topic of the talk. As soon as I heard him say this [Option D], I knew D was the answer before the question arrived.*

Similarly, 14 (82.4%) of the 17 test-naïve students employed the prediction strategy and got the correct answer based on the first statement of the speaker. However, the remaining 3 (17.6%) test-naïve students did not read ahead the options nor predict the question. Rather, they just concentrated on listening and selected their answers (1 correct) based on their understanding of the talk.

Item 43 in Listening Comprehension

Of the 23 test-wise students, 19 (82.6%) claimed to have read at least two options prior to listening but none of them were able to predict the question. One test-wise student said:

> *I quickly glanced through the first two options because I ran out of time. I couldn't tell what they were going to ask...perhaps...perhaps something about what they did? I don't know. So, I skipped this one and continued to read the next set of options.*

Fourteen (60.9%) test-wise students reported that they heard the speaker saying something like "They traveled aimlessly...far from it!" Based on this information, these

students selected A, which was opposite of the word "aimlessly". One test-wise student described how he arrived at the answer:

> *I barely heard something like "traveled aimlessly" and then "far from that." I guessed that the opposite of it means that they traveled with an aim. So, I chose A.*

One (4.3%) test-wise student picked Option C, which contained the word "cave" she heard in the talk (IIB4). Three (13.0%) additional students relied on the information they obtained to eliminate 3 options as incorrect (ID1): one successfully eliminated the 3 distracters; and the other two first ruled out C and D, then A, using the obtained information. One test-wise student explained how he eliminated options:

> *I didn't catch some details. Anyway, C is dead wrong. What did they ask? It seemed to me that they asked why those people didn't live in the cave. The cave was found by themselves...but not for hiding...well...anyway, it is dead wrong. D is not mentioned in the talk. The answer seemed to be between A and B. I heard that they moved towards any place where they could find food. That means that they didn't plan their migrations. So, it has to be B.*

The remaining 5 (21.7%) test-wise students claimed that they knew the answer (all correct).

Twelve (70.6%) of the 17 test-naïve students read ahead at least two options and none of them were able to figure out what the question would be. Six (35.3%) test-naïve students picked A claiming that they heard the speaker saying something like "they planed their moves" (IIB4). Three (17.6%) additional students selected their answers (all incorrect) based on the information they believed they heard from the talk. Their responses, respectively, were:

> *Student A:* *I heard two sentences. One was "...not far." The other was "They could not go too far away because they put lots of things in the cave." So, B is correct.*

Student B: *I heard that they found the cave and hid themselves in it. I selected C.*

Student C: *I reasoned that, since they stored their food and supplied in the cave, they wouldn't travel far. I chose B.*

One (5.9%) test-naïve student successfully ruled out the 3 foils using the information obtained from the talk (ID1). The remaining 7 (41.2%) test-naïve students claimed that they knew the answer (6 correct).

In sum, items 36, 42, and 43, while identified as not susceptible to test-wisenss, somehow became "construct-irrelevant-easy" (Messick, 1989) when the Chinese subsample students, particularly test-wise students, employed some of the test-taking skills to respond these items. The empirical evidence obtained from the protocol analysis of the students' "think-aloud" data showed that the students performance on these items was somewhat influenced by what test-wise strategies were applied and how these strategies were used. For example, item 36 was the most difficult item in the TOEFL Interview Form for the subsample of the Chinese students (see Table 11). When the students failed to find the correct answer using their knowledge alone, higher proportions of the use of test-taking strategies and "educated" guessing were observed for both the test-wise and test-naïve students (see Table 12). Nevertheless, in this case, quality beats quantity. Although the stem-option connection (IIB4) was the most common strategy employed by both test-wise and test-naïve subgroups, 14 (60.9%) test-wise students were able to catch the key word "triangle" from the conversation and then connected this word with Option C, the correct answer, whereas only 3 (17.6%) test-naïve students managed to do the same. This difference in the use of the IIB4 strategy resulted in a 40.2% difference in mean scores between these two subgroups (see Table 11).

Item 42 seemed to be vulnerable to the "prediction" and stem-option (IIB4) strategies. The vast majority (92.5%) of the 40 students in the interview predicted the up-coming question before listening and, consequently, caught the key phrase which matched Option C, the correct answer, from the very first sentence of the speaker. In contrast, the 3 test-naïve students who did not predict the question failed to correctly respond to this item.

A successful response to item 43 depended on whether a student was able to catch and comprehend the key phrase "far from it" in the talk. In this respect, compared to the test-naïve students, the test-wise subgroup illustrated their stronger ability to catch the key word/phrase in the talk and link it with an option (IIB4). While the proportions to utilize the IIB4 strategy was close (i.e., 60.9% vs. 52.9%) between the test-wise and test-naïve subgroups, the test-wise students were more successful obtaining the right information and, consequently, responding to the item correctly. Overall, they outperformed the test-naïve students by 10.5% in mean scores on item 43.

Third Category

Items 45, 46, and 48 in Reading Comprehension belong to the third category. These items were basically "clean," with no judgmental or empirical evidence to show the item's susceptibility to test-wiseness.

Items 45, 46, and 48 in Reading Comprehension

The three items were all based on one reading passage about the problems after the American Civil War. After reading the text, students were asked to select the best answer from among the 4 options provided.

(45) The passage refers to all of the following as necessary steps following the Civil War EXCEPT

(A) helping soldiers readjust
(B) restructuring industry
(C) returning government to normal
*(D) increasing taxes

(46) The word "task" in line 15 refers to
(A) raising the tax level
(B) sensible financial choices
(C) wise decisions about former salves
*(D) reconstruction of damaged areas

(48) The word "them" in line 26 refers to

(A) charges
*(B) leaders
(C) days
(D) irons

Note: * the keyed option

Item 45 in Reading Comprehension

All of the 23 test-wise students selected Option D, the correct answer. Two (8.7%) test-wise students chose D based on their "common sense" that "it was absurd to increase taxes after the war." One (4.3%) student eliminated the 3 distracters (ID1). He said:

> *A, B, and C were all mentioned in the text. D was absurd. It is impossible to increase the taxes after the war. Otherwise, people would suffer more. According to what I understood about the text and my common sense, it [D] was not right. I didn't find this [D], anyway.*

The remaining 20 (87.0%) test-wise students claimed that they knew the answer.

Of the 17 test-naïve students, all but one (94.1%) correctly responded to item 45, claiming that they knew the answer. One student selected A because she didn't find the word "help" in the text. She explained:

I didn't think A was right. They only said that soldiers had to readjust to civilian life but never said that the government would help them. At least, I didn't find it. Obviously, A was the answer and I didn't need to look further.

Items 46 and 48 in Reading Comprehension

The 23 test-wise and 17 test-naïve students correctly responded to these two items. They all claimed that they knew the answer. There was no evidence to suggest that they applied any of test-taking strategies when they responded to these two items.

In sum, the results of the protocol analysis of the students' "think-aloud" data revealed that the 40 students, both test-wise and test-naïve, basically did not employ any test-wiseness strategy when responding to the 3 non-susceptible items. No discernible difference in test-taking behaviors was found between the test-wise and test-naïve subgroups. It was observed that, with only one exception, all of the students were able to locate the correct answer based on their content knowledge alone. In this circumstance, the use of test-wiseness strategies seemed to become unnecessary.

Summary of Chinese Students' Test-wiseness Behaviors Based on the TOEFL Interview Data

First, the differences between the test-wise and test-naïve subgroups can be seen in Table 12. In many cases, the primary strategies used by the test-wise students were similar to or in the trend with the target strategies identified by the two judges and confirmed by the empirical evidence for the full sample. Further, it appears that the differences between test-wise and test-naïve subgroups in terms of the patterns of solution strategies used sometimes may accounted completely or partly for the difference in group mean scores for a TOEFL item. Evidence to support this point was found in the

students' responses to items 41 and 49 in Reading Comprehension. Item 41 was identified as susceptible to the convergence strategy (ID4). When responding to this item, on the one hand, 15 (65.2%) of the 23 test-wise students applied the convergence strategy alone (4 students), combined ID4 with ID1 (9 students), or used ID1 alone (2 students) to figure out the correct answer whereas none of the 17 test-naïve students used the convergence strategy and only 3 (17.6%) of them utilized ID1. On the other hand, more test-naïve students (82.4%) claimed that they knew the answer than the test-wise students (43.5%). As a result, the test-wise students outperformed the test-naïve students by 13.3% with respect to the proportions of students in each group who correctly answered the item. For item 49, differences between test-wise and test-naïve groups were found in terms of the patterns of the solution strategies used by each group. Compared to the test-wise students, none of the test-naïve students identified Option A and C as opposites, fewer used the ID1 and/or IIB4 strategies, and more claimed that they knew the answer. Consequently, the mean percentage (p = 95.7%) for the test-wise group was considerably higher than the mean percentage (p = 47.1%) for the test-naïve group. In these two cases, the inference could be made that the differences between the test-wise and test-naïve subgroups in the mean scores on items 41 and 49 may be attributable to the differences between the two subgroups in the patterns of solution strategies applied in responding to the questions. This finding may shed some light on understanding how the Chinese sample students approached the TOEFL items using various test-wiseness strategies.

Second, regardless of what test-wise elements the items were susceptible to, elimination of options known to be incorrect was the most-used strategy for both of the test-wise and test-naïve subgroups. Students in general tended to employ the ID1 strategy alone or in conjunction with other test-wise strategies. Also, the test-wise students tended to use the ID1 for double purposes: (1) to figure out the correct answer by eliminating an option known to be incorrect when they did not know the answer or were not quite sure about the answer; and (2) to confirm/double check whether their selection was the best when they felt that they knew the answer.

Third, the second most commonly-used strategy, particularly when responding to TOEFL listening items, was to establish a stem-option connection (IIB4) by employing a key phrase/word, scattered information, a vague general idea, and the speaker's tone obtained from the conversation/talk provided. In the case when a key word in the conversation/talk was unknown or was not obtained (e.g., item 27 in Listening Comprehension), the prior knowledge or what was called "common sense" was used as partial knowledge to link the stem with an option. Using this kind of the strategy allowed students to attain the correct answer without necessarily gaining a thorough understanding of the conversation/talk provided.

Fourth, "educated guessing" was found in both the test-wise and test-naïve students' responses (e.g., items 27, 37, and 44 in Listening Comprehension, and items 42, 44, 47, and 49 in Reading Comprehension). As defined in Chapter I (see p.5), "educated guessing" refers to a specific test-taking behavior that students first eliminate one or more options as incorrect and then guess randomly from among the remaining options. This strategy, although not the best solution, is better than random guessing and allows

students to increase the probability of getting the correct answer for an item. For example, when responding to item 27 in Listening Comprehension, none of the students except one knew the meaning of the word "pushover." In spite of this, six of the 40 students attempted the question using test-wiseness strategies first to eliminate one or two options and then to guess from among the remaining options. Of them, four test-wise students ruled out two options using either the ID3 strategy (2 students) or the ID1 strategy (2 students), and then guessed randomly from the remaining two options (2 correctly). Two test-naïve students also eliminated two options as incorrect (ID1), and then randomly guessed from the remaining two options (both correctly).

Fifth, it seems that the differences between the test-wise and test-naïve subgroups not only exist in the frequency of using the test-wise strategies but also in the quality and effectiveness of the application of these strategies. In many cases (e.g., items 37 and 44 in Listening Comprehension, and items 42, and 44 in Reading Comprehension), it was quality rather than quantity of the application of test-wiseness skills that accounted for the difference in the group mean scores between the two subgroups. Compared to the test-naïve students, the test-wise students' approaches seemed to be more meaningful, thoughtful, logical, defendable, and less random. In the case when they tried to answer the question using their prior knowledge or "common sense," the test-wise students in general appeared to be more academically knowledgeable than their test-naïve counterparts. In addition, the test-wise students seemed to be more cautious about their selection based on their knowledge alone and less frequently claimed that they knew the answer than the test-naïve students.

Last, it appears that the effective application of test-wise strategies relies heavily upon partial knowledge. This partial knowledge, although inadequate to respond to the TOEFL items, when combined with the application of test-wise strategies, increases the probability of correctly responding to items containing test-wise cues. Consequently, the total test score will be inflated. If this partial knowledge is considered relevant to the construct of the TOEFL, then interpretation of the total test score as a valid indicator of language proficiency may be justified. Yet to the extent this partial knowledge is considered irrelevant to what the TOEFL test is designed to measure, the interpretation of the test score will be confounded by construct-irrelevant easiness (Messick, 1989) and the validity of inferences made from test scores will be under the threat.

CHAPTER VI

SUMMARY, CONCLUSIONS, AND IMPLIATIONS FOR PRACTICE AND FUTURE RESEARCH

Chapter VI consists of four sections. The first section is a brief summary of the first three chapters. The second section contains a summary of findings from Chapters IV and V. The limitations, implications, and suggestions for practice and future research are presented and discussed in the last two sections.

Overall Summary of the Study

Test-wiseness is defined as "a subject's capacity to utilize the characteristics and formats of the test and/or test-taking situation to receive a high score" (Millman, Bishop, & Ebel, 1965, p.707). If an examinee possesses partial knowledge as well as test-wiseness and if the test contains susceptible items, then the combination of these factors could inflate the test score (Rogers & Bateson, 1991a).

The Test of English as a Foreign Language (TOEFL) is the largest and the most influential English test in the world. It provides scores obtained from multiple-choice items that contribute to decisions regarding admission to or exclusion from more than 2, 400 colleges and universities in the United States and Canada. Nevertheless, is the TOEFL immune to test-wiseness? If not, what is influence of test-wiseness upon performance on the TOEFL? To date, there has been little research to answer these questions.

Therefore, the purpose of the present study was to examine the effects of test-wiseness upon performance on the TOEFL. In order to achieve this objective, methods

involving both quantitative and qualitative approaches were adopted. First, a sample of 390 Chinese TOEFL candidates from a TOEFL advance training program at Shanghai Qianjin Institute in China was selected. These students were asked to respond to a modified version of the Test of Test-wiseness (TTW; see p. 59) and then, following a ten-minute recess, wrote the TOEFL Practice Test B (ETS, 1995d), to which they had not been previously exposed. Based on the results of item analysis of the TTW, a subsample of 40 students, consisting of the 23 test-wise and 17 test-naive students, was identified and then interviewed on an individual basis. In the interview, each student was asked to "think aloud" about the strategies he/she was using while responding to each item in the Interview Form which contained the test-wiseness susceptible items of the TTW and TOEFL. The interviews were audio-taped, and then transcribed and translated into English.

The presence of test-wise susceptible items in the TOEFL was first identified through consensus judgement done by the two independent judges and then verified by the statistics of item analysis of the TOEFL based on the entire sample (n = 390). Items were not considered to be susceptible to test-wiseness unless identified by both the judgmental and empirical processes.

In order to understand what general cognitive processes the Chinese sample students usually employed when responding to test-wise susceptible items in the TTW and TOEFL, protocol analysis of the interview data was conducted. Comparison was made between test-wise and test-naïve students in terms of the strategies used when responding to the test-wise susceptible items of the TTW and the TOEFL.

Summary of the Findings

Findings Related to the TTW

1. Based on the results of the item analysis of the TTW, the majority (80.8%) of the Chinese sample students (n = 390) involved in this study possessed some test-wiseness abilities. Further, it may be concluded that approximately 70% or more of the sample students were able to identify and use the cues related to incorrect options (ID1), similar options (ID2), and opposite options (ID3). Nonetheless, given that only 41.0% of the students scored above the upper chance level on the IIB4 subtest, the students as a group appeared to be less aware of stem-option similarities (IIB4). More pronounced, the Chinese students seemed to be much less capable of recognizing specific determiners (IIB3); 78.7% of the students scored at or below the chance level on this subscale, a finding consistent with what was reported by Lo and Slakter (1973) and Wu and Slakter (1978).

2. In light of the results of "think-aloud" data analysis, the prime strategies used by the test-wise students in many cases were the same or similar to the target strategies the TTW was designed to measure. However, given the unique background and outlooks in culture, education, curricula, social values, ideology and political systems, the Chinese students, when searching for an answer, did not always employ the target strategies as expected. In addition, incompetence in English language may jeopardize their ability to recognize and correctly use test-

wise cues such as stem-option connection (IIB4), similar options (ID2), and opposite options (ID3).

3. The ID1 strategy appeared to be the most frequently used strategy by both the test-wise and test-naïve groups. Students in general tended to apply this strategy either alone or in combination with other test-wiseness skills. The second most-used strategy was to find a link between the stem and an option (IIB4). Overall, it seemed to be a common approach among the sample students to utilize their partial knowledge to eliminate options known to be incorrect (ID1) and/or to identify a connection between the stem and an option (IIB4). In this aspect, the students classified as test-wise seemed to be more academically knowledgeable and, consequently, more successful in application of the ID1 and IIB4 strategies than the students classified as test-naïve. Nevertheless, as a whole group, the Chinese students' approach was not always effective in approaching the TTW items where either the language was too difficult for them to comprehend or the entire content was unfamiliar to them.

4. The test-wise students were more successful than the test-naïve students in recognizing and making use of similar options (ID2) and opposite options (ID3). The test-naïve students had difficulty in not only identifying but also appropriately using the ID2 or ID3 strategy. On the one hand, when confronted with the ID2 items, they tended to choose between the two similar options. On the other hand, when facing a pair of options with subtle differences (ID3), they

were likely to avoid both options even though they were aware that the two options were not exactly the same in meaning.

5. The subsample of students (n = 40) in the interviews did not do well on items containing specific determiners (IIB3), either. Their poor performance may be attributed to the nature of the training program they were involved. As presented in Chapter V, the TOEFL test was free of IIB3 cues and, consequently, students didn't have to learn how to recognize and use IIB3 cues. In addition to the students' insufficient knowledge about specific determiners, the lack of American cultural and historical background may be another factor to account for the low mean scores on the IIB3 items. Due to their unique historical and political background, over a half of the subsample students found nothing wrong with the specific determiners like "all", "every" and "never" when these words appeared in the context of law, regulation, and legal/political claims in an extreme period such as wartime.

6. In general, compared with the test-naïve students, test-wise students' approaches appeared to be more thoughtful, logical, appropriate, defendable, rational, and less random. This difference may be attributed to two underlying factors. First, the test-wise students seemed to be more academically knowledgeable than the test-naïve students. Second, the test-wise students appeared to be more persistent in search for test-wiseness cues. They seldom randomly guessed or picked an option as the correct answer only because it was the shortest/longest (IIB1a) or read

better grammatically (IIB1h). Instead, they kept looking for subtler and multiple cues until they were sure that there were no more cues or a satisfactory answer had been obtained. This finding fits Rogers and Bateson's model of test-wise test taking behaviour (see Figure 1, p. 3) as well as their observations about the differences between test-wise and test-naïve students (1991a, 1991b).

Findings Related to the TOEFL

1. Based on the evidence obtained from the item analysis of the TOEFL (n = 390), 48% to 64% of the items across the Listening and Reading Comprehension Sections of the TOEFL Practice Test B (ETS, 1995d) were identified as susceptible to test-wiseness. According to the results of correlated t-test, the means for the items identified as susceptible to test-wiseness exceeded significantly ($p < 0.01$) the corresponding means for the non-testwiseness items in both the Listening and Reading Comprehension Sections.

2. The most common test-wise cue found in the two sections of the TOEFL Practice Test B (ETS 1995d) considered was absurd options (ID1), followed by opposite options (ID3). The number of items containing ID2 (similar options), IIB4 (stem-option similarity), ID4 (convergence), and ID5 (the correct answer embedded in other test items) was, respectively, 1, 3, 1, and 3. No items containing specific determiners (IIB3) was found from the two sections considered.

3. In many cases, the prime strategies employed by the test-wise students were similar to, or in the trend with the target strategies identified by the two judges and confirmed by the empirical evidence gathered from the full sample. Nevertheless, the strategy most frequently used by both the test-wise and test-naïve subgroups was elimination of options known to be incorrect (ID1). Students in general tended to utilize the ID1 strategy alone or in conjunction with other test-wise strategies. For test-wise students, application of ID1 often served double purposes: (1) to figure out the correct answer by eliminating options known to be incorrect when they did not know, or were not sure about the answer; and (2) to double check their final selection when they felt that they knew the answer.

4. The second most commonly-used strategy, particularly when responding to the TOEFL listening items, was to establish stem-option connection using partial knowledge (IIB4). The partial knowledge here often included a key phrase/word, scattered information, a vague general idea, and the speaker's tone obtained from the conversation/talk provided. In the case when a key word in the conversation/talk was unknown or was not obtained (e.g., item 27 in Listening Comprehension), the prior knowledge or what was called "common sense" was utilized as partial knowledge to link the stem with an option. Using this kind of the strategy allowed students to figure out the correct answer without necessarily gaining a thorough understanding of the conversation/talk provided. Again, the

test-wise students outperformed the test-naïve students in the successful application of the IIB4 strategy.

5. "Educated guessing" was found from both the test-wise and test-naïve students' responses to some of the listening and reading TOEFL items (e.g., items 27, 37, and 44 in Listening Comprehension, and items 42, 44, 47, and 49 in Reading Comprehension). As defined in Chapter I (see p. 7), "educated guessing" refers to a specific test-taking behavior that students first eliminate one or more options as incorrect and then guess randomly from among the remaining options. This strategy, although not the best solution, is better than randomly guessing and allows students to increase the probability of getting the correct answer for an item.

6. One prevailing strategy applied by the interview subsample students in responding to the TOEFL listening items was to predict questions and the content/topic of the talk/conversation using the 4 options provided before listening to the tape. All of the students in the interviews admitted that this strategy was one of major test-taking strategies they were trained to learn in class. They cited that the "prediction" strategy could help them roughly know beforehand what topic/theme of the talk and up-coming question might be so that they could listen selectively for relevant information and key words. Differences between the test-wise and test-naïve students were found in the use of the "prediction" strategy. In general, 82.6.0% to 100% of the 23 test-wise students

applied or claimed to apply the "prediction" strategy when responding to each of the TOEFL Interview listening items whereas 64.7% to 88.2% of the 17 test-naïve students reported that they did the same. Further, compared to the test-naïve students, the test-wise students not only more frequently applied the "prediction" strategy but also were more capable of using this strategy effectively. It appeared that, in many cases (e.g., items 11, 27, 34, 42, 45 in Listening Comprehension), the effective use of the "prediction" strategy did help the test-wise students ease the test anxiety and obtain some relevant information before listening and, consequently, enhanced their performance on the listening subtest of the TOEFL.

7. As indicated earlier, differences between the test-wise and test-naïve subgroups not only existed in the frequency of using the test-wise strategies but also in the quality and effectiveness of the application of these strategies. In a few cases (e.g., items 49 and 41 in Reading Comprehension), the difference between the two subgroups in performance on the TOEFL items may be accounted completely or partly by the different patterns of solution strategies each subgroup applied. Yet in many cases (e.g., items 37 and 44 in Listening Comprehension, and items 42, and 44 in Reading Comprehension), the difference in performance seemed to be attributable to the difference between the two subgroups in quality and effectiveness of using test-wiseness skills. Compared with the test-naïve students, the test-wise students' approaches seemed to be more meaningful, thoughtful, logical, defendable, and less random. In the case when they tried to answer the question using their prior knowledge or "common sense," the test-wise students in

general appeared to be more academic knowledgeable than their test-naïve counterparts. In addition, the test-wise students seemed to be more cautious about their selection based on their knowledge alone and less frequently claimed that they knew the answer than the test-naïve students.

8. It appeared that the effective application of test-wise strategies relies heavily upon partial knowledge. This partial knowledge, although inadequate to respond to the TOEFL items solely on the basis of fully understanding the reading texts, the conversations/talks, and the questions provided, was sufficient, when combined with the application of test-wise strategies, to increase the probability of correctly responding to items containing test-wise cues. Given this finding, the total test score would be inflated. If this partial knowledge is considered relevant to the construct of the TOEFL when enhanced by a student's capacity to utilize clues present in items, then interpretation of the total test score as a valid indicator of language proficiency may be justified. Yet to the extent this partial knowledge is considered irrelevant to what the TOEFL test was designed to measure, the interpretation of the test score would be confounded by construct-irrelevant easiness (Messick, 1989) and the validity of inferences made from test scores would be under the threat.

Limitations

Presented below are limitations of the study.

1. The external validity and generalizability of the research results was limited by the nature of the sample involved. The subjects involved were restricted to Chinese adults in a TOEFL training program in one city in China. Given the complex nature of human demographic factors, such as cultural/language background, residency areas (urban vs. rural), age, and educational level, the findings from this investigation may not necessarily be generalizable to other Chinese students nor to the whole population of the TOEFL candidates across cultures and nations. By the same token, the findings of this study may not be necessarily generalizable to other forms of the TOEFL.

2. Given that the major target test-wiseness skills (i.e., ID1, ID2, ID3, and IIB4) on which the present study was focussed were mainly based upon Millman et al's (1965) Taxonomy of Test-wiseness Principles (pp. 711-712), other test-wiseness strategies not listed in the Taxonomy but commonly used on the TOEFL by the students may not have been adequately assessed. Consequently, the differences in test-wiseness between the test-wise and test-naïve students when responding to the TOEFL items may not have been adequately addressed.

3. An interval of 10 days between the administration of the TOEFL and the interviews may not have been long enough for the subsample students to forget the content of the tests initially administered. It was found that some students

still remembered the content of some items. Students in general, by listening to the tape the second time in the interview, performed better on the TOEFL listening items than they did on the same items at the first time. A test-retest effect might have confounded some of the results about the subsample in the interviews.

4. The selection of non-testwise items was somewhat problematic and may cause misinterpretation. Only three of the eight items were truly non-susceptible to test-wiseness. Of the remaining five items, two were identified as susceptible to test-wiseness by the judges but not supported by empirical evidence; and three were identified by the judges as non-susceptible but were found to be vulnerable to a certain test-wiseness strategy. Given that these five so-called "non-testwise" items were not really free of test-wiseness cues, the interpretation of the lower score on these five items for the test-naïve group was confounded with effects of test-wiseness and the levels of language proficiency the students possessed.

5. Due to the lack of an independent and non-testwise measure of proficiency, whether the levels of proficiency across the test-wise and test-naive groups were the same is unknown. The interpretation of differences in performance between the test-wise and test-naive groups may have been confounded with the possible differences in the levels of proficiency between the two groups. There may also have been a confounding of test-wiseness with metalinguistic awareness or language aptitude. Without an independent and non-testwise measure, it is

impossible to isolate the effect of metalinguistic awareness in a test that requires close attention to linguistic cues.

Implications for Practice

It is important to reiterate that influence of test-wiseness upon performance on the TOEFL is a concern with respect to not only the validity of test score interpretations but also to fair student assessment. According to the current use of TOEFL scores in the tertiary admission process in North America, students are compared based on their total TOEFL scores. If TOEFL scores can be influenced by test-wiseness as evidence obtained from the study suggests, then the difference between high and low scores would be spuriously larger than if only English language proficiency, or the lack of thereof, was contributing to this difference. In this case, students well trained in test-wiseness application would benefit by taking advantage of item clues whereas students who are unaware of test-wiseness would be penalized, even though the general level of English proficiency of the students is equivalent. Thus for fairness and equity, something must be done to minimize the "differential" effects of test-wiseness. On one hand, both the test administration process and test format need to be improved to minimize the susceptibility to test-wiseness. One approach to improve item quality, as Rogers and Harley (1999) recommended, would be to reduce the number of options in a multiple-choice item from four to three. Compared with the four-option format, the three-option format, while maintaining the internal consistency reliability unchanged, will reduce the challenge for test-developers to construct multiple-choice items all of which contain the required number of plausible, equally attractive foils to students who do not know the answer

based on subject knowledge and understanding alone. On the other hand, as in many cases, if it is not possible to eliminate test-wiseness completely, according to Standard 11.13 *of Standards for Educational and Psychological Testing* (AREA, 1999), test-wiseness instruction materials and information should be provided to all TOEFL candidates as part of a test preparation guidance so that all of the TOEFL candidates across the world would be made aware of test-wiseness. If all TOEFL candidates have learned test-wiseness, then the effects of test-wiseness upon performance would relatively become a constant.

Implications for Future Research

1. The results of the present study revealed that mere knowledge of the concept of test-wiseness strategies does not guarantee improvement in performance on tests. The students who participated in this research had all had test-taking strategy instruction, but clearly some made better use of these strategies than others. This difference, while partially attributable to the variable length and intensity of the test-wiseness instruction each student received, was largely accounted for by the students' different levels of linguistic and background knowledge and their ability to use test-wiseness. The effective application of test-wiseness skills depends on partial knowledge of both English and subject matter as well as the ability to apply these skills.
2. It was observed in this study that the students' facility with their second language played a crucial role in the application of test-wiseness strategies. Test-wiseness is a skill that can be used to take tests, including language tests, but the ability to

apply test-wiseness strategies on an examination is contingent upon a certain degree of language proficiency and aptitude. Given that the subject matter being assessed was language proficiency itself, the interpretation of the test score and student performance may have been confounded with the overlap of test-wiseness, language proficiency and metalinguistic awareness. What are the inter-relationships among these three factors? Is it possible to separate them? If the content knowledge being assessed is language proficiency, can test-wiseness still be considered to be "logically independent of the examinee's knowledge of the subject matter for which the items are supposedly measures" (Millman et al., 1965, p. 707)? To answer these questions, further research is needed. One approach to exploring test-wiseness holding language proficiency constant would be to replicate the research using educated English native speakers. Alternatively, a study of test-wiseness in which language proficiency and metalinguistic awareness are partialled out using ANCOVA may also elucidate the effect of test-wiseness and establish the degree to which test-wiseness alone contributes to performance on a proficiency test such as the TOEFL. Research that isolates test-wiseness may indicate inequities faced by students who do not have access to test-wiseness training.

3. It was observed that the students utilized some listening strategies (e.g., the "prediction" strategy) to help them respond to the TOEFL listening items. These strategies, while strongly advocated in TOEFL test preparation materials/instruction in particular, are not included in Millman et al.'s (1965)

Taxonomy of Test-wiseness Principles. Given that tests are increasingly administered using multimedia technology, corresponding test-taking strategies will inevitably develop. It is suggested that further research be conducted to expand the taxonomy of test-wiseness. In addition, since TOEFL tests are currently administered by computer in most parts of the world, a study is needed to identify and verify whether there are test-wiseness strategies unique to the computer format.

4. The use of internal consistency estimates to measure the reliability of test-wiseness instruments may not be appropriate and meaningful. The low internal consistency of the TTW was consistently reported in many studies (Slakter, 1970; Rogers & Bateson, 1991a; Vanchu, 1990; Man, 1990). This low internal consistency may be attributed to student's random guessing and particularly an "educated" guessing, and, therefore, is inevitable. Thus the current methods of assessing the reliability of test-wiseness instruments may be inadequate and inaccurate. Using other measures of stability such as point-biserial coefficients may be more appropriate in assessment of test-wiseness skills.

References

Abadzi, J. (1976). Evaluation of Foreign students [sic] admission procedures used at the University of Alabama (Doctoral dissertation, University of Alabama, 1975). Dissertation Abstracts International, 36, 7754A. (University Microfilm No. 76-13, 884)

Alderman, D. L., & Holland, P. W. (1981). Item performance across native language groups on the Test of English as a Foreign Language (TOEFL Research Rep. No. 9; ETS Research Rep. No. 81-16). Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 218 922)

Allen, A. (1992). Development and validation of a scale to measure test-wiseness in EFL/ESL reading test takers. Language Testing, 9, 101-122.

American Educational Research Association. (1999). Standards for educational and psychological testing. Washington, DC: Author.

Anastasi, A. (1976). Psychological testing (4th ed.) New York: Macmillan.

Angelis, P. J. (1977). Language Testing and Intelligence Testing: Friends or foes? In J. E. Reddon (Ed.), Proceedings of the first International Conference on Frontiers in Language Proficiency and Dominance Testing. Occasional Papers on Linguistics, No. 1. Carbondale, IL: Southern Illinois University. (ERIC Document Reproduction Service No. ED 145 677)

Angoff, W. H. (1989). Does guessing really help? Journal of Educational Measurement, 26 (4), 323-336.

Angoff, W. H., & Sharon, A. T. (1971). A comparison of scores earned on the Test of English as a Foreign Language by native American college students and foreign applicants to US colleges. TESOL Quarterly, 5, 129-36.

Ardiff, M. B. (1965). The relation of three aspects test-wiseness to intelligence and reading ability in grades three and six. Unpublished master thesis, Cornell University, New York.

Ayrer, J., Diamond, J. J., Fishman, R., & Green, P. (1976). Are inner city children test-wise? Journal of Educational Measurement, 14, 39-45.

Ayers, J. B., & Peters, R. M. (1977). Predictive validity of the Test of English as a Foreign Language for Asian graduate students in engineering, chemistry, or mathematics. Educational and Psychological Measurement, 37, 461-463.

Ayers, J. B., & Quattlebaum, R. F. (1992). TOEFL performance and success in masters program in engineering. Educational and Psychological Measurement, 52, 973-975.

Bachman, (1990). Fundamental considerations in language testing. Oxford: Oxford University Press.

Bajtelsmit, J. W. (1975, October). Development and validation of an adult measure of secondary cue-using strategies on objective examinations: The test of obscure knowledge (TOOK). Paper presented at the annual meeting of the Northeastern Educational Research Association, Ellenville, NY.

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C-L. C. (1983). Effects of coaching programs on achievement test performance. Review of Educational Research, 53, 571-585.

Benson, J. (1985, April). Distinguishing test-wiseness strategies among ethnically diverse elementary students. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Benson, J. (1988). The psychometric and cognitive aspects of test-wiseness: A review of the literature. In M. H. Kean (Ed.), Test-wiseness (pp. 1-25). Bloomington, IN: Phi Delta Kappan.

Benson, J., Urman, H., & Hocevar, D. (1986). Effects of test-wiseness training and ethnicity on achievement and third-and fifth-grade students. Measurement and Evaluation in Counselling and Development, 18, 154-162.

Birenhaum, M., & Nasser, F. (1994). On the relationship between test anxiety and test performance. Measurement & Evaluation in Counselling & Development, 27, 293-301.

Borrello, G. M., & Thompson, B. (1985). Correlates of selected test-wiseness skills. Journal of Experimental Education, 53, 124-128.

Brown, A. L. (1980). Metocognitive development and reading. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension (pp. 453-481). Hillsdale, NJ: Erlbaum.

Brown A. L. (1987). Metacognition executive control, self-regulation and other even more mysterious mechanisms. In F. E. Wienert & R. H. Kluwe (Eds.), Metacognition, motivation, and understanding (pp. 65-116). Hillsdale, NJ: Erlbaum.

Brown, J. D. (1993). Differential subtest performance and test-taker characteristics? University of Hawaii Working Papers in ESL, 12 (1), 139-169.

Buell, J. G. (1992, March). TOEFL and IELTS as measures of Academic reading Ability: An exploratory study. Paper presented at the annual meeting of the Teachers of English to Speakers of Other Languages. (ERIC Document Reproduction Service No. ED 354 758)

Burgess, T. G., & Greis, N. B. (1970). English language proficiency and academic achievement among students of English as a second language at the college level. (ERIC Document Reproduction Service No. ED 074 812)

Callenbach, C. (1973). The effects of instruction and practice in content-independent test-taking techniques upon the standardized reading test scores of selected second grade students. Journal of Educational Measurement, 10, 25-30.

Carman, R. A. & Adams, W. R. (1972). Study skills: A student's guide for survival. New York: Wiley.

Carter, K. (1986). Test-wiseness for teachers and students. Educational Measurement: Issues and practice, 5 (4), 20-23.

Choy, S. C., & Davenport, B. M. (1986). The TOEFL: Incomplete test of English Proficiency. College Teaching, 34, 108-110.

Clark, J. L. D. (1977). The performance of native speakers of English on the Test of English as a Foreign Language (TOEFL Research Rep. No. 1). Princeton, NJ: Educational Testing Service.

Clark, J. L. D., & Swinton, S. S. (1980). The Test of Spoken English as a measure of communicative ability in English-medium instructional settings (TOEFL Research Rep. No. 7; ETS Research Rep. No. 80-33). Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 218 960)

Crehan, K. D., Koehler, R. A., & Slakter, M. J. (1974). Longitudinal studies of test-wiseness. Journal of Educational Measurement, 11, 209-212.

Crehan, K. D., Gross, L. J., & Slakter, M. J. (1978). Developmental aspects of test-wiseness. Educational Research Quarterly, 3 (1), 40-44.

CTB/McGraw Hill. (1974). Comprehensive Test of Basic Skills - Form S, Level 6.5-8.9. Monterey, CA: Author.

Des Brisay, M., & Ready, D. (1991). Defining an appropriate role for language tests in intensive English language programs. In S. Anivan (Ed.), Issues in Language programme Evaluation in the 1990s (pp. 15-25). Singapore: SEAMEO Regional Language Centre.

Diamond, J. J., & Evans, W. J. (1972). An investigation of the cognitive correlates of test-wiseness. Journal of Educational Measurement, 9, 145-150.

Diamond, J. J., Ayres, J., Fishman, R., & Green, P. (1976). Are inner city children test-wise? Journal of Educational Measurement, 14, 39-45.

Dizney, H. (1965). Concurrent validity of the Test of English as a Foreign Language for a group of foreign students at an American university. Educational and Psychological Measurement, 25, 1129-1131.

Durost, W. W., Bixler, H. H., J., Wrightstone, J. W., Prescott, G. A., & Balow, I. (1970). Metropolitan Achievement Test - Form F, Intermediate Level. New York: Harcourt, Brace and Jovanovich, Inc.

Educational Testing Service. (1968). TOEFL test manual. Princeton, NJ: Author.

Educational Testing Service. (1973). TOEFL test manual. Princeton, NJ: Author.

Educational Testing Service. (1983). TOEFL test manual. Princeton, NJ: Author.

Educational Testing Service. (1990). Bulletin of Information for TOEFL and TSE, 1990-91. Princeton, NJ: Author.

Educational Testing Service. (1991a). Bulletin of Information for TOEFL and TSE, 1991-92. Princeton, NJ: Author.

Educational Testing Service. (1991b). TOEFL--2000: Planning for change. Princeton, NJ: Author.

Educational Testing Service. (1992). Bulletin of Information for TOEFL/TWE and TSE, 1992-93. Princeton, NJ: Author.

Educational Testing Service. (1995a). The Researcher. Princeton, NJ: Author.

Educational Testing Service. (1995b). Bulletin of Information for TOEFL and TSE, 1995-96. Princeton, NJ: Author.

Educational Testing Service. (1995c). TOEFL test manual. Princeton, NJ: Author.

Educational Testing Service. (1995d). TOEFL Practice Tests. Princeton, NJ: Author.

Educational Testing Service. (1997). Overview. TOEFL test & score manual. Princeton, NJ: Author.

Educational Testing Service. (2000). TOEFL 2000-2001 Information Bulletin for Computer-Based Testing. Princeton, NJ: Author.

Ely, M., Anzul, M., Friedman, T., Garner, D., & Steinmetz, A. M. (1991). Doing qualitative research: Circles within circles. London, England: Falmer Press.

Erickson, M. E. (1972). Test sophistication: An important consideration. Journal of Reading, 16, 140-144.

Ericsson, K. A., & Simon (1993). Protocol analysis (Rev. ed.). Cambridge, MA: The MIT Press.

Flavell, J. H. (1979). Metacognition and cognitive monitoring. American Psychologist, 34, 906-911.

Flippo, F. (1987). Test Wise: Strategies for success in taking tests. Belmont, CA: David S. Lake Publishers.

Farhady, H. (1982). Measures of language proficiency from the learner's perspective. TESOL Quarterly, 16, 43-59.

Ford, V. A. (1973). Everything you want to know about test-wiseness. Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 093 912)

Gaines, W. G., & Longsma, E. A. (1974, April). The effect of training in test-taking skills on the achievement scores of fifth grade pupils. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago.

Geranpayeh, A. (1994). Are score comparisons across language proficiency test batteries justified?: An IELTS-TOEFL comparability study. Edinburgh Working Papers in Applied Linguistics, 5, 50-65.

Gibb, B. C. (1964). Test-wiseness as a secondary cue response. Unpublished doctoral dissertation, Stanford University, CA.

Glass, G. V, & Hopkins, K. D. (1997). Statistical methods in education and psychology (7th ed.). Englewood Cliffs, NJ: Prentice-Hall, INc.

Graham, J. G. (1987). English language proficiency and the prediction of Academic success. TESOL Quarterly, 21, 505-521.

Gross, L. J. (1976, April). The effects of three selected aspects of test-wiseness on the standardized test performance of eighth grade students. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.

Gue, L. R., & Holdaway, E. A. (1973). English proficiency Tests as predictors of success in graduate studies in education. Language Learning, 23, 89-103.

Gullickson, A. (1986). Teacher education and teacher preservice needs in educational assessment and evaluation. Journal of Educational Measurement, 22, 327-354.

Gullickson, A., & Hopkins, K. (1987). The context of educational measurement instruction for preservice teachers: Professor perspectives. Educational Measurement: Issues & practice, 6 (3), 12-16.

Hale, G. A., Stansfield, C. W., & Duran, R. P. (1984). Summaries of studies involving the Test of English as a Foreign Language, 1963-1982 (TOEFL Research Rep. No. 16). Princeton, NJ: Educational Testing Service.

Hale, G. A. (1988). The interaction of student major-field group and text content in TOEFL Reading Comprehension (TOEFL Research Rep. No. 25). Princeton, NJ: Educational Testing Service.

Harmon, M. G., Morse, D. T., & Morse, L. W. (1994, November). Confirmatory factor analysis of the Gibb Experimental test of testwiseness. Paper presented at the Mid-South Educational Research Association, Nashville, TN.

Heil, D. K., & Aleamoni, L. M. (1974). Assessment of the proficiency in the use and understanding of English by foreign students as measured by the Test of English

as a Foreign Language (Report RR-350). Urbana: Univ. of Illinois. (ERIC Document Reproduction Service No. ED 093 948)

Ho, D. Y. F., & Spinks, J. A. (1985). Multivariate prediction of academic performance by Hong Kong University students. Contemporary Educational Psychology, 10, 249-259.

Hosley, D., & Meredith, K. (1979). Inter- and intra-test correlates of the TOEFL. TESOL Quarterly, 13, 209-217.

Hoyt, C. (1941). Test reliability estimated by analysis of variance. Psychometrika, 6, 153-160.

Hughes, A. (1989). Testing for language teachers. Cambridge: Cambridge University Press.

Hwang, K., & Dizney, H. F. (1970). Predictive validity of the Test of English as a Foreign Language for Chinese graduate students at an American university. Educational and Psychological Measurement, 30, 475-477.

Irvine, P., Atai, P., & Oller, J. W., Jr. (1974). Cloze, dictation and Test of English as a Foreign Language. Language Learning, 24, 245-252.

Jameson, D. C., & Malcolm, D. J. (1973). TOEFL--The developing years. International Educational and Cultural Exchange, 8, 57-62.

Jiang, H. (1999). Establishment of score distributions for TOEFL Concordance Tables. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Montreal, Quebec, Canada. (ERIC Document Reproduction Service No. ED 431 808)

Johnson, D. C. (1977). The TOEFL and domestic students: Conclusively inappropriate. TESOL Quarterly, 11, 79-86.

Joint Advisory Committee. (1993). The principles for fair student assessment practices for Education in Canada. Edmonton, Alberta: Author.

Kippendorff, K. (1980). Content analysis: An introduction to its methodology. Beverly Hills, CA: Sage.

Kulik, C. L., & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. American Educational Research Journal, 19 (3), 415-428.

Light, R. L., Xu, M., & Mossop, J. (1987). English proficiency and academic performance of international students. TESOL Quarterly, 21 (2), 251-262.

Lo, M., & Slakter, M. J. (1973). Risk-taking and test-wiseness of Chinese students. The journal of Experimental Education, 42 (2), 56-59.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Man, D. W. (1990). Cross-cultural study of test-wiseness. Unpublished Master thesis. The University of British Columbia, Vancouver, BC.

Mehrens, W. A., & Lehmann, I. J. (1973). Measurement and evaluation in Education and Psychology. New York: Holt, Rinehart, and Winston, Inc.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational Measurement (pp. 13-103). Washington, DC: American Council on Education and Macmillan Publishing Company.

Metfessel, N. S., & Sax, G. (1958). Systematic biases in the keying of correct responses on certain standardized tests. Educational and Psychological Measurement, 18, 787-790.

Miller, P. M., Fuqua, D. R., & Fagley, N. S. (1990). Factor structure of the Gibb experimental test of testwiseness. Educational and Psychological Measurement, 50, 203-208.

Millman, J. (1966). Test-wiseness in taking objective achievement and aptitude examinations (Final report). New York: College Entrance Examination Board.

Millman, J., Bishop, H., & Ebel, R. (1965). An analysis of test-wiseness. Educational and Psychological Measurement, 25, 707-726.

Millman, J., & Setijadi, A. (1966). A comparison of the performance of American and Indonesian students on three types of test items. Journal of Educational Research, 59, 273-275.

Morse, D. T. (1994, November). The relative difficulty of selected test-wiseness skills among college students. Paper presented at the annual meeting of the Mid-South Educational Research Association, Nashville, NT (ERIC Document Reproduction Service No. 387 528).

Mullen, K. A. (1978). Determining the effect of uncontrolled sources of error in a direct test of oral proficiency and the capability of the procedure to detect improvement following classroom instruction. In J. L. D. Clark (Ed.), Direct testing of speaking proficiency: Theory and application. Princeton, NJ: Educational Testing Service.

Oakland, T. (1972). The effects of test-wiseness material on standardized test performance of preschool disadvantaged children. Journal of School Psychology, 10, 355-360.

Odunze, O. J. (1982). Test of English as Foreign Language and first year GPA of Nigerian students. Dissertation Abstracts International, 42, 3419A-3420A.

Oller, J. W., Jr. & Spolsky, B. (1979). The test of English as a Foreign Language. In B. Spolsky (Ed.), Some major tests: Advanced in Language Testing Series, (pp. 92-100). Arlington, VA: Centre for Applied Linguistics. (ERIC Document Reproduction Service No. 183 004)

Oltman, P. K., Stricher, L. J., & Barrows, T. (1988). Native Language, English proficiency, and the structure of the Test of English as Foreign Language (TOEFL Research Rep. No. 27). Princeton, NJ: Educational Testing Service.

Pack, A. C. (1972). A comparison between TOEFL and Michigan test scores and student success in (1) freshman English and (2) completing a college program. TESL Reporter, 5, 1-7, 9.

Palmer, L. A. (1965). TOEFL: Testing of English as a Foreign Language. Bulletin of the National Catholic Educational Association, 62, 235-238.

Palmer, I. C. (1984). The ethic of test preparation at intensive English language programs. (ERIC Document Reproduction Service No. ED 248727)

Pierce, B. N. (1992). Demystifying the TOEFL reading test. TESOL Quarterly, 26, 665-689.

Pike, L. W. (1979). An evaluation of alternative item formats for testing English as a foreign language (TOEFL Research Rep. No. 2). Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 206 627)

Prell, J. M., & Prell, P. A. (1986). Improving test scores--teaching test-wiseness--A review of the literature. Phi Delta Kappa, 5, 2-5.

Powers, D. E., & Leung, S. W. (1995). Answer the new SAT reading comprehension questions without the passages. Journal of Educational Measurement, 32 (2), 105-129.

Pryczak, F. (1973, February). Use of similarities between stem and keyed choice in multiple choice items. Paper presented at the annual meeting of the National Council of Measurement in Education, New Orleans.

Raimes, A. (1990). The TOEFL Test of Written English: Cause for concern. TESOL Quarterly, 24 (3), 427-442.

Rogers, W. T. (1991). Educational assessment in Canada: Evolution or extinction? Alberta Journal of Educational Research, 37 (2), 79-92.

Rogers, W. T., & Bateson, D. J. (1990). The impact of test-wiseness upon year-end school leaving examinations. Manuscript submitted for publication.

Rogers, W. T., & Bateson, D. J. (1991a). The influence of test-wiseness on performance of high school seniors on school leaving examinations. Applied Measurement in Education, 4, 159-183.

Rogers, W. T., & Bateson, D. J. (1991b). Verification of a model of test-taking behaviour of high school seniors. Journal of Experimental Education, 54, 331-350.

Rogers, W. T., & Wilson, C. (1993). The influence of test-wiseness upon performance of high school students on Alberta Education's Diploma Examinations (Work Report). Unpublished paper.

Rogers, W. T., & Yang, P. (1996). Testwiseness: Its nature and application. European Journal of Psychological Assessment. Tiburg, Netherland.

Rogers, W. T., & Harley, D. (1999). An empirical comparison of three- and four-choice items and tests: Susceptibility to testwiseness and internal consistency reliability. Educational and Psychological Measurement, 59 (2), 234-247.

Rowley, G. L. (1974). Which examinees are most favoured by the use of multiple choice tests? Journal of Educational Measurement, 11, 15-23.

Sarnacki, R. E. (1979). An examination of test-wiseness in the cognitive test domain. Review of Educational Research, 49, 252-279.

Sax, G., & Carr, A. (1962). An investigation of response sets on altered parallel forms. Educational and Psychological Measurement, 22, 371-376.

Scholz, G. E., & Scholz, C. M. (1981, Detroit). Multiple-choice cloze tests of EST discourse: An exploration. Paper presented at the fifteenth annual TESOL convention. (ERIC Document Reproduction Service No. ED 208 656)

Schuell, T. J. (11986). Cognitive conceptions of learning. Review of Educational Research, 56, 411-436.

Scruggs, T. E., & Lifson, S. A. (1985). Current conceptions of test-wiseness: Myths and realities. School Psychology Review, 14 (3), 339-350.

Scruggs, T. E. et al. (1985). Learning disabled students' spontaneous use of test-taking skills on reading achievement tests. Learning Disability Quarterly, 8 (3), 205-210.

Shanghai Progressing Institute. (1992). A brief introduction to Shanghai Progressing Institute [Brochure]. Shanghai: Author.

Sharon, A. T. (1972). English proficiency, verbal aptitude, and foreign student success in American graduate schools. Educational and Psychological Measurement, 32, 425-431.

Slack, W. D., & Porter, D. (1980). The SAT: A critical appraisal. Harvard Educational Review, 50, 154-175.

Slakter, M. J. (1969). Generality of risk taking on objective examinations. Educational Psychological Measurement, 29 (1), 115-128.

Slakter, M. J. Koehler, R. A., & Hampton, S. H. (1970). Grade level, sex, and selected aspects of test-wiseness. Journal of Educational Measurement, 7, 119-122.

Smith, J. K. (1980). The convergence strategy of test-wiseness. Paper presented at the annual meeting of the American Research Association, MA: Boston.

Smith, J. K. (1982). Converging on correct answers: A peculiarity of multiple-choice items. Journal of Educational Measurement, 19, 211-219.

Spolsky, B. (1990). The prehistory of TOEFL. Language Testing, 7, 98-118.

Swinton, S. S., & Power, D. F. (1980). Factor analysis of the Test of English as a Foreign Language for several language groups (TOEFL Research Rep. No. 6; ETS Research Rep. No. 80-32). Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 218 921)

Thorndike, E. L. (1951). Reliability. In E. F. Lindguist (Ed.), Educational measurement (pp. 560-620). Washington, DC: American Council on Education.

Traynor, R. (1985). The TOEFL: An appraisal. ELT Journal, 39, 43-47.

Urman, H. (1983). The effect of test-wiseness training on the achievement of third and fifth grade students. Paper presented at the annual meeting of the National Council on Measurement in Education. Montreal, Quebec.

Vernon, P. (1962). The determinants of reading comprehension. Educational and Psychological Measurement, 22, 269-286.

Warner, C. J. (1982). A study of the relationships between TOEFL scores and selected performance characteristics of Arabic students in an individualized competency based training program (Doctoral dissertation, University of Tennessee, Knoxville, 1981). Dissertation Abstracts International, 42, 4810A. (University Microfilms No. DA 8209009)

Weiten, W. (1984). Violation of selected item construction principles in educational measurement. Journal of Experimental Education, 52, 174-178.

Wilcox, L. O. (1975). The prediction of Academic success of undergraduate foreign students. Dissertation Abstract International, 35, 6084B.

Wilson, K. M. (1982). A comparative analysis of TOEFL examinee characteristics, 1977-1979 (TOEFL Research Rep. No. 11, ETS Research Rep. No. 82-27). Princeton, NJ: Educational Testing Service.

Wilson, K. M. (1987). Patterns of test taking and score change for examinees who repeat the Test of English as a Foreign Language (TOEFL Research Rep. No. 22). Princeton, NJ: Educational Testing Service.

Woodford, P. E. (1978). English tests: Their credibility in foreign student admission. College and University, 54, 500-507.

Woodley, K. K. (1973, April). Test-wiseness program development and evaluation. Paper presented at the annual meeting of the American Educational Research Association. New Orleans.

Wu, T. H., & Slakter, M. J. (1978). Risk-taking and test-wiseness of Chinese students by grade level and residence area. Journal of Educational Research, 71 (3), 167-170.

Yalden, J. (1978). TOEFL and the management of foreign students enrollments. TESL Talk, 9, 16-21.

APPENDIX A

TEST OF TEST-WISENESS

BACKGROUND INFORMATION

(Please fill it in Chinese)

Name: _____

Language School:_____

Age:_____

Birth Place: _____

Sex M/F

Highest Degree: _____

Year of Graduation and University:

_____

Major: _____

Your Address (where you can be contacted):

_____

_____

_____

_____

Post Code: _____

Telephone: _____

How long have you studied English? _____

Have you ever had any coaching or specific lessons on how to take the TOEFL test?

0. No, never.
1. Yes, once or twice.
2. Yes, three or more times.

Have you ever seen or written the form of the TOEFL test used in this research or some of the questions from this test form?

No/Yes. If yes, when did you see/write this form of the TOEFL last time?

_____

Have you ever written the TOEFL in the regular test center?

No/Yes. If yes, how many times? _____. And what is your latest TOEFL score (if applicable)? _____.

Name:_____ School:_____

TEST OF TEST-WISENESS

This is a test of test-wiseness which measures some of the abilities needed to do well on tests. Many of the questions are about things you may not have studied. However, there are test-taking strategies which can be used to figure out what to do when faced with such questions.

For example:

The greatest advantage of using slent in the manufacture of steel is that slent makes steel

A. transparent.
B. stainless.
C. heavy.
D. rubbery.

Using test-wiseness strategies, Option 'A' and 'D' can be eliminated since they are clearly not correct (steel is not transparent, nor is it rubbery). Therefore, either 'B' or 'C' is the correct answer. Now we stand a better chance of guessing the correct answer for we have narrowed the number possible options down to two from four.

INSTRUCTIONS: For each question, select the BEST answer and record your choice on the answer sheet provided. Each question is worth one mark. There will be no correction for guessing.

1. Compared to normal cells, bileuvial cells

A. divide more rapidly.
B. divide more slowly.
C. have more cytoplasm.
D. have more mitochondria.

2. The Flying Spider is known for its abilities to

A. blend in with its surrounding.
B. glide through the air.
C. kill its prey with poison.
D. make very large webs.

3. **The square root of 1.1 can be best approximated by**
   A. the cube root of 11056.
   B. the solution of $x^2 - 16 = 0$.
   C. using the binomial series.
   D. factoring the expression $9x^2 - 18$.

4. **Hermann Klavemann is best known for**
   A. developing all musical scales used in the western world.
   B. composing every sonata during the romantic era.
   C. translating all Russian classics into English.
   D. inventing the safety pin.

5. **Mr. Adams, in Henry Fielding's Joseph Andrews,**
   A. learns his parents were of the nobility.
   B. takes sick after falling through the ice.
   C. falls into the mud while reading.
   D. discovers he is of noble birth.

6. **The Proclamation of 1763**
   A. forbade colonists to settle territory acquired in the French and Indian wars.
   B. encouraged colonists to settle territory acquired in the French and Indian wars.
   C. provided financial incentives for settlement of territory acquired in the French and Indian wars.
   D. all of the above.

7. **Which of the following would help to determine if D is the fourth harmonic of C with respect to A and B?**
   A. The relative size of angle ACB to angle ADB.
   B. The length if line segment AB.
   C. The fact that A, B, C, and D lie on one straight line.
   D. The straight line distance from A to B.

8. **A normal percentage of polymorphonuclear leukocytes found in the human peripheral blood is**
   A. 53/260.
   B. 70%.
   C. 115%.
   D. 035.

9. **In the Dartmouth College case the United States Supreme Court held that**
   A. the court had no right under any circumstances ever to nullify an Act of Congress.
   B. a state could not impair a contract.
   C. all contracts must be agreeable to the state legislature.
   D. all contracts must inevitably be certified.

10. Wordsworth's "The Prelude" (1805)

A. tells of a descent into Hell in a Model-T Ford.
B. makes use of a distinction between "the sublime" and "the beautiful".
C. is concerned with the emerging African nations.
D. was influenced by Hemingway's **The Sun Also Rises**.

11. The literature of the early eighteenth century is

A. public in nature, relating to society's outlooks and values.
B. private in nature, relating to an individual's emotions and feelings.
C. rough and irregular compared to the literature of the later eighteenth century.
D. filled with despair over the apparent collapse of traditional values.

12. Charles Dickens' Hard Times deals with

A. the difficult life of a factory worker.
B. The politics of the French chateau country.
C. The court of King Edward III.
D. The limitations of European existentialism.

13. Organisms of the Pavo genus

A. change from masculine to feminine gender.
B. display their plumage for the female.
C. become female after existing for a period of time as male.
D. possess an excess quantity of masculine hormone.

14. Which of the following most likely caused the War of 1693?

A. Spain was building roads to connect her cities.
B. France was going through great agricultural change.
C. France believed that Spain was increasing her troops.
D. Spain had a series of earthquakes.

15. A spherical triangle is the triangle on the surface of a sphere. What name is given to the number of degrees in a spherical triangle minus 180?

A. The arc of the triangle.
B. The size if the triangle.
C. The spherical excess of the triangle.
D. The polar measurement of the triangle.

16. In Horace Walpole's The Castle of Otranto

A. Manfred is the father of Hippolita.
B. Hippolita is the wife of Manfred.
C. Manfred is the uncle of Hipplita.
D. Hippolita is the daughter of Manfred.

17. The ring F[x]/s(x) is a field if and if only

A. s(x) is a prime polynomial over F.
B. s(x) is a rational polynomial over F.
C. F(x) is a multiple of s(x).
D. any element if F(x) contains an inverse.

18. The career of Marius (155-86 B.C.), the opponent of Sulla, is significant in Roman history because he

A. gave many outstanding dinners and entertainments for royalty.
B. succeeded in arming the gladiators.
C. showed that the civil authority could be trusted aside by the military.
D. made it possible for the popular party to conduct party rallies outside the city of Rome.

19. A substance that, in its pure form, is the best conductor of electricity is

A. water.
B. deuterium.
C. H_2O.
D. silver.

20. The august character of the work Pericles in Athens frequently causes his work to be likened to that in Rome of

A. Augustus.
B. Sulla.
C. Pompey.
D. Claudius.

21. "Lucifer in Starlight" is a

A. modern psychological story of World War II.
B. Shakespearean sonnet by Gerard Manley Hopkins.
C. controversial French novel written by Resnails.
D. Petrarchan sonnet by George Meredith.

22. The treaty of Brest Litovsk was ratified by Moscow because

A. Tsar Alexander I wanted to prevent Napoleon's invasion of Russia.
B. Russia was unable to keep up with the armament manufacture if Austria.
C. Russia could not keep pace with the military production of Austria.
D. Nicolai Lenin wanted to get the Soviet Union out of World War I.

23. How many iambic feet (one iambic foot-one unstressed syllable followed by one stressed syllable, as in "perFORM") are in each line of Robert Pack's poem "The Compact"?

A. 1.
B. 5
C. 16
D. 22.

24. What is the probability that a needle of length L<D, when dropped on a table ruled with equidistant parallel lines of distance D apart, will cross one of the lines?

A. $\frac{\sqrt{3}}{3}(LD)$

B. $\frac{2}{\pi D}$

C. L + .001D

D. $\frac{2D}{\pi L}$

25. The Feulgen Nuclear Reaction demonstrates the presence of

A. desoxyribonucleoptrotein.
B. lysosomes.
C. mitochondria.
D. endoplasmic reticulum.

26. The Alabama claims were

A. all settled completely and satisfactorily.
B. claims against Jefferson Davis for seizure of all of the property in the state during wartime.
C. claims of the United States against Great Britain.
D. claims of every citizen of Alabama against every citizen of Georgia.

APPENDIX B

INTERVIEW FORM

Name:_____ School:_____

TEST OF TEST-WISENESS

This is a test of test-wiseness which measures some of the abilities needed to do well on tests. Many of the questions are about things you may not have studied. However, there are test-taking strategies which can be used to figure out what to do when faced with such questions.

For example:

The greatest advantage of using slent in the manufacture of steel is that slent makes steel

E. transparent.
F. stainless.
G. heavy.
H. rubbery.

Using test-wiseness strategies, Option 'A' and 'D' can be eliminated since they are clearly not correct (steel is not transparent, nor is it rubbery). Therefore, either 'B' or 'C' is the correct answer. Now we stand a better chance of guessing the correct answer for we have narrowed the number possible options down to two from four.

INSTRUCTIONS: For each question, select the **BEST** answer and tell me orally your choice in terms of A, B, C, and D. Also, please explain how you have arrived at the answer. Please be advised that the following items are selected from Test of Test-wiseness and they are not necessarily in sequential order.

Warm-up Exercise:

The flying spider is known for its ability to

A. blend in with its surrounding.
B. glide through the air.
C. kill its prey with poison.
D. make very large webs.

4. Hermann Klavemann is best known for

A. developing all musical scales used in the western world.
B. composing every sonata during the romantic era.
C. translating all Russian classics into English.
D. inventing the safety pin.

5. Mr. Adams, in Henry Fielding's Joseph Andrews,

A. learns his parents were of the nobility.
B. takes sick after falling through the ice.
C. falls into the mud wile reading.
D. discovers he is of noble birth.

6. The Proclamation of 1763

A. forbade colonists to settle territory acquired in the French and Indian wars.
B. encouraged colonists to settle territory acquired in the French and Indian wars.
C. provided financial incentives for settlement of territory acquired in the French and Indian wars.
D. all of the above.

9. In the Dartmouth College case the United States Supreme Court held that

A. the court had no right under any circumstances ever to nullify an Act of Congress.
B. a state could not impair a contract.
C. all contracts must be agreeable to the state legislature.
D. all contracts must inevitably be certified.

10. Wordsworth's "The Prelude" (1805)

A. tells of a descent into Hell in a Model-T Ford.
B. makes use of a distinction between "the sublime" and "the beautiful".
C. is concerned with the emerging African nations.
D. was influenced by Hemingway's **The Sun Also Rises**.

11. The literature if the early eighteenth century is

A. public in nature, relating to society's outlooks and values.
B. private in nature, relating to an individual's emotions and feelings.
C. rough and irregular compared to the literature of the later eighteenth century.
D. filled with despair over the apparent collapse of traditional values.

21. "Lucifer in Starlight" is a

A. modern psychological story of World War II.
B. Shakespearean sonnet by Gerard Manley Hopkins.
C. controversial French novel written by Resnails.
D. Petrarchan sonnet by George Meredith.

22. The treaty of Brest Litovsk was ratified by Moscow because

A. Tsar Alexander I wanted to prevent Napoleon's invasion of Russia.
B. Russia was unable to keep up with the armament manufacture if Austria.
C. Russia could not keep pace with the military production of Austria.
D. Nicolai Lenin wanted to get the Soviet Union out of World War I.

23. How many iambic feet (one iambic foot-one unstressed syllable followed by one stressed syllable, as in "perFORM") are in each line of Robert Pack's poem "The Compact"?

A. 1.
B. 5
C. 16
D. 22.

24. What is the probability that a needle of length L<D, when dropped on a table ruled with equidistant parallel lines of distance D apart, will cross one of the lines?

A. $\frac{\sqrt{3}}{3}(LD)$

B. $\frac{2}{\pi D}$

C. L + .001D

D. $\frac{2D}{\pi L}$

26. The Alabama claims were

A. all settled completely and satisfactorily.
B. claims against Jefferson Davis for seizure of all of the property in the state during wartime.
C. claims of the United States against Great Britain.
D. claims of every citizen of Alabama against every citizen of Georgia.

TEST OF ENGLISH AS A FOREIGN LANGUAGE

Scenario: A new and inexperienced TOEFL test taker comes to you for your advice about the best way to approach the following items selected from a TOEFL test. Please help him by answering these items orally in terms of A, B, C, or D and explaining to him how you have arrived at the answer.

Warm-up Exercise:

Listening Comprehension

1. (A) He makes a lot of money.
   (B) He has just been left some money.
   (C) He doesn't believe three hundred dollars is enough.
   (D) He can't afford to spend that much.

Note:

Page 212 has been removed due to copy right restrictions. The information removed was Items 10, 11, 27, 34, 35, 36, and 37 excerpted from Section 1: Listening Comprehension, TOEFL Practice Test B (ETS, 1995d).

Note:

Page 213 has been removed due to copy right restrictions. The information removed was Items 42, 43, 44, 45, and 46 excerpted from Section 1: Listening Comprehension, TOEFL Practice Test B (ETS, 1995d).

Note:

Page 214 has been removed due to copy right restrictions. The information removed was a reading passage excerpted from Section 3: Reading Comprehension, TOEFL Practice Test B (ETS, 1995d).

Note:

Page 215 has been removed due to copy right restrictions. The information removed was Items 41, 42, 43, 44, 45, 46, and 47 excerpted from Section 3: Reading Comprehension, TOEFL Practice Test B (ETS, 1995d).

Note:

Page 216 has been removed due to copy right restrictions. The information removed was Items 48, 49, and 50 excerpted from Section 3: Reading Comprehension, TOEFL Practice Test B (ETS, 1995d).

APPENDIX C

SCRIPT TO GUIDE THE INTERVIEW

INTRIDUCTION

Suppose I am a new and inexperienced TOEFL test-taker. I am coming to you for your advice on how to take the TOEFL test. I am going to ask you to respond on the following items selected from the TOEFL and from the Test of Test-Wiseness. I would like you to answer each item orally and explain how you have arrived at your answer. To me, it doesn't matter whether your answer is right or wrong. I am more interested in your process that leads to your answer. Therefore, please tell me exactly what you are doing and thinking when you responding to each item. Your name, test score and any personal information will not be revealed to anyone. Our conversation in the interview will be used anonymously.

PROCEDURES

1. At the beginning of each interview, approximately 2 to 3 minutes will be set aside for an informal conversation in order to ease the anxiety of the interviewee and to facilitate the communication between the interviewee and the interviewer.

2. To reduce the unnecessary variation in the process of data collection, the interviewer will strictly follow the script that has been prepared to guide each interview. The interviewer will make sure that each interviewee understands the purpose of the study and the procedures to be used.

3. The interviewee will be provided with a practice item on each of the two subtests (i.e., the TTW and TOEFL) of the Interview Form. After the interviewee becomes familiar with the think-aloud procedure, the interview will officially start and recorder will be on.

4. During the interview, probes will be minimized and only used to get at the sequence of behaviors and special strategies that are employed. For example, instead of asking "Did you read the option first?" the interviewer will ask, "What was first thing you did when you approach this item?" To reduce the interviewee's anxiety and suspicion, the interviewer will not take any field notes. Also, the interviewer will remain as silent as possible to minimize unnecessary interruptions. If the interviewee pauses for a while, the interviewer will ask the interviewee, "What are you thinking?" or simply say, "Keep talking ," or "Go on."

APPENDIX D

RESULTS OF ITEM ANALYSIS FOR THE TTW

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 13

TEST NO 1 TTW.SH SUBTEST 1 ID1 SUBTEST

ITEM NUMBER 3

OPTION	WT	N	P		COEFFICIENTS OF CORRELATION PB-ST	PB-TT	B-ST	B-TT		MEANS ST	TT
1	0	40	10.3		-.28	-.18	-.47	-.31		2.08	11.68
2	0	31	7.9		-.19	-.15	-.35	-.28		2.26	11.71
C 3	1	261	66.9	C	.49	.25	.63	.33	C	3.41	13.79
4	0	55	14.1		-.27	-.06	-.42	-.10		2.25	12.80
OTHER	0	3	.8		.00	-.02	-.01	-.07		3.00	12.67
TOTAL		390									

ITEM NUMBER 8

OPTION	WT	N	P		COEFFICIENTS OF CORRELATION PB-ST	PB-TT	B-ST	B-TT		MEANS ST	TT
1	0	71	18.2		-.33	-.22	-.48	-.32		2.21	11.89
C 2	1	269	69.0	C	.42	.25	.55	.33	C	3.34	13.76
3	0	13	3.3		-.15	-.11	-.37	-.26		2.08	11.54
4	0	36	9.2		-.13	-.05	-.22	-.10		2.56	12.75
OTHER	0	1	.3		-.04	.05	-.29	.30		2.00	16.00
TOTAL		390									

ITEM NUMBER 10

OPTION	WT	N	P		COEFFICIENTS OF CORRELATION PB-ST	PB-TT	B-ST	B-TT		MEANS ST	TT
1	0	48	12.3		-.14	-.04	-.23	-.06		2.58	12.94
C 2	1	73	18.7	C	.31	.17	.45	.25	C	3.77	14.33
3	0	121	31.0		.02	.05	.02	.07		3.05	13.48
4	0	147	37.7		-.16	-.15	-.21	-.19		2.78	12.68
OTHER	0	1	.3		-.09	-.09	-.57	-.57		1.00	8.00
TOTAL		390									

ITEM NUMBER 14

OPTION	WT	N	P		COEFFICIENTS OF CORRELATION PB-ST	PB-TT	B-ST	B-TT		MEANS ST	TT
1	0	20	5.1		-.21	-.28	-.43	-.58		2.00	9.70
2	0	56	14.4		-.34	-.17	-.52	-.26		2.07	12.04
C 3	1	298	76.4	C	.46	.32	.63	.44	C	3.31	13.79
4	0	14	3.6		-.17	-.09	-.40	-.20		2.00	11.93
OTHER	0	2	.5		.00	.01	-.01	.03		3.00	13.50
TOTAL		390									

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 14

TEST NO 1 TTW.SH SUBTEST 1 ID1 SUBTEST

ITEM NUMBER 18

OPTION	WT	N	P		COEFFICIENTS OF CORRELATION PB-ST	PB-TT	B-ST	B-TT		MEANS ST	TT
1	0	17	4.4		-.06	-.04	-.13	-.08		2.71	12.76
2	0	115	29.5		-.17	-.07	-.22	-.10		2.72	12.92
C 3	1	141	36.2	C	.44	.25	.56	.32	C	3.69	14.26
4	0	116	29.7		-.26	-.17	-.35	-.22		2.55	12.48
OTHER	0	1	.3		-.04	-.07	-.29	-.46		2.00	9.00
TOTAL		390									

ITEM NUMBER 23

OPTION	WT	N	P		COEFFICIENTS OF CORRELATION PB-ST	PB-TT	B-ST	B-TT		MEANS ST	TT
1	0	74	19.0		-.14	-.03	-.21	-.04		2.68	13.07
C 2	1	135	34.6	C	.43	.19	.55	.24	C	3.70	14.03
3	0	138	35.4		-.24	-.10	-.31	-.13		2.64	12.84
4	0	38	9.7		-.08	-.07	-.14	-.12		2.74	12.61
OTHER	0	5	1.3		-.08	-.06	-.28	-.22		2.20	11.60
TOTAL		390									

1 LERTAP 2.0 SUBTEST STATISTICS
PAGE 15

TEST NO 1 TTW.SH SUBTEST 1 ID1 SUBTEST

NUMBER OF INDIVIDUALS	= 390.00	NUMBER OF ITEMS	= 6.00
MEAN	= 3.02	HIGHEST SCORE	= 6.00
STANDARD DEVIATION	= 1.15	LOWEST SCORE	= .00

SOURCE OF VARIANCE	D.F.	S.S.	M.S.
INDIVIDUALS	389.00	86.48	.22
ITEMS	5.00	107.26	21.45
RESIDUAL	1945.00	391.24	.20
TOTAL	2339.00	584.98	.25

HOYT ESTIMATE OF RELIABILITY = .10

STANDARD ERROR OF MEASUREMENT = 1.00

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 16

TEST NO 1 TTW.SH SUBTEST 2 ID2 SUBTEST

ITEM NUMBER 5

OPTION	WT	N	P		COEFFICIENTS OF CORRELATION PB-ST	PB-TT	B-ST	B-TT		MEANS ST	TT
1	0	45	11.5		-.19	-.12	-.32	-.20		3.36	12.27
C 2	1	72	18.5	C	.32	.20	.46	.29	C	4.82	14.51
C 3	1	82	21.0	C	.25	.04	.35	.06	C	4.60	13.51
4	0	189	48.5		-.33	-.12	-.41	-.15		3.60	12.90
OTHER	0	2	.5		.00	.01	.00	.03		4.00	13.50
TOTAL		390									

ITEM NUMBER 7

OPTION	WT	N	P		COEFFICIENTS OF CORRELATION PB-ST	PB-TT	B-ST	B-TT		MEANS ST	TT
C 1	1	138	35.4	C	.15	.03	.19	.04	C	4.25	13.39
2	0	33	8.5		-.25	-.20	-.45	-.36		3.00	11.30
C 3	1	166	42.6	C	.18	.16	.23	.21	C	4.26	13.82
4	0	41	10.5		-.22	-.10	-.38	-.17		3.22	12.37
OTHER	0	12	3.1		-.12	-.06	-.31	-.14		3.17	12.33
TOTAL		390									

ITEM NUMBER 13

OPTION	WT	N	P		COEFFICIENTS OF CORRELATION PB-ST	PB-TT	B-ST	B-TT		MEANS ST	TT
1	0	61	15.6		-.22	-.11	-.34	-.17		3.38	12.49
C 2	1	82	21.0	C	.09	.00	.13	.01	C	4.22	13.28
3	0	79	20.3		-.28	-.22	-.40	-.31		3.33	11.96
C 4	1	157	40.3	C	.38	.31	.48	.39	C	4.57	14.38
OTHER	0	11	2.8		-.17	-.16	-.43	-.40		2.82	10.55
TOTAL		390									

ITEM NUMBER 16

OPTION	WT	N	P		COEFFICIENTS OF CORRELATION PB-ST	PB-TT	B-ST	B-TT		MEANS ST	TT
1	0	47	12.1		-.26	-.15	-.42	-.24		3.17	12.09
C 2	1	154	39.5	C	.26	.20	.33	.25	C	4.40	13.97
C 3	1	126	32.3	C	.16	.05	.21	.06	C	4.29	13.46
4	0	60	15.4		-.31	-.18	-.48	-.28		3.12	11.98
OTHER	0	3	.8		-.10	-.05	-.40	-.19		2.67	11.67
TOTAL		390									

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 17

TEST NO 1 TTW.SH SUBTEST 2 ID2 SUBTEST

ITEM NUMBER 19 COEFFICIENTS OF CORRELATION MEANS

	OPTION	WT	N	P		PB-ST	PB-TT	B-ST	B-TT		ST	TT
	1	0	34	8.7		-.26	-.17	-.46	-.30		3.00	11.65
C	2	1	68	17.4	C	.06	.02	.09	.04	C	4.16	13.41
	3	0	29	7.4		-.32	-.26	-.59	-.49		2.66	10.52
C	4	1	259	66.4	C	.28	.23	.37	.29	C	4.25	13.73
	TOTAL		390									

ITEM NUMBER 22 COEFFICIENTS OF CORRELATION MEANS

	OPTION	WT	N	P		PB-ST	PB-TT	B-ST	B-TT		ST	TT
C	1	1	133	34.1	C	.22	.11	.29	.14	C	4.38	13.71
	2	0	58	14.9		-.34	-.25	-.52	-.38		3.02	11.48
	3	0	71	18.2		-.29	-.16	-.43	-.23		3.25	12.27
C	4	1	126	32.3	C	.29	.22	.38	.29	C	4.52	14.21
	OTHER	0	2	.5		-.12	-.08	-.57	-.38		1.00	10.00
	TOTAL		390									

1 LERTAP 2.0 SUBTEST STATISTICS
PAGE 18

TEST NO 1 TTW.SH SUBTEST 2 ID2 SUBTEST

NUMBER OF INDIVIDUALS = 390.00 NUMBER OF ITEMS = 6.00

MEAN = 4.01 HIGHEST SCORE = 6.00

STANDARD DEVIATION = 1.21 LOWEST SCORE = .00

SOURCE OF VARIANCE	D.F.	S.S.	M.S.
INDIVIDUALS	389.00	95.50	.25
ITEMS	5.00	47.44	9.49
RESIDUAL	1945.00	376.06	.19
TOTAL	2339.00	519.00	.22

HOYT ESTIMATE OF RELIABILITY = .21

STANDARD ERROR OF MEASUREMENT = .98

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 19

TEST NO 1 TTW.SH SUBTEST 3 ID3 SUBTEST

ITEM NUMBER 1 COEFFICIENTS OF CORRELATION MEANS

	OPTION	WT	N	P		PB-ST	PB-TT	B-ST	B-TT		ST	TT
C	1	1	176	45.1	C	.32	.24	.41	.31	C	1.57	14.05
C	2	1	37	9.5	C	.06	-.08	.10	-.15	C	1.35	12.49
	3	0	88	22.6		-.21	-.10	-.29	-.14		2.66	12.69
	4	0	88	22.6		-.22	-.13	-.30	-.18		2.64	12.56
	OTHER	0	1	.3		-.05	.00	-.30	-.03		2.00	13.00
	TOTAL		390									

ITEM NUMBER 6 COEFFICIENTS OF CORRELATION MEANS

	OPTION	WT	N	P		PB-ST	PB-TT	B-ST	B-TT		ST	TT
C	1	1	100	25.6	C	.25	.26	.34	.16	C	1.65	14.59
C	2	1	91	23.3	C	.29	.05	.40	.07	C	3.77	13.54
	3	0	58	14.9		-.23	-.08	-.35	-.13		2.47	12.67
	4	0	138	35.4		-.29	-.21	-.37	-.27		2.64	12.41
	OTHER	0	3	.8		-.13	-.09	-.53	-.36		1.33	10.33
	TOTAL		390									

ITEM NUMBER 11 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	B-ST	B-TT		ST	TT
C 1	1	76	19.5	C	.21	.01	.30	.01	C	3.64	13.29
C 2	1	138	35.4	C	.24	.04	.31	.06	C	3.53	13.43
3	0	105	26.9		-.28	-.10	-.38	-.14		2.55	12.74
4	0	67	17.2		-.17	.05	-.25	.08		2.67	13.60
OTHER	0	4	1.0		-.07	.03	-.27	.13		2.25	14.25
TOTAL		390									

ITEM NUMBER 17 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	B-ST	B-TT		ST	TT
C 1	1	64	16.4	C	.22	.11	.33	.16	C	3.73	13.97
C 2	1	101	25.9	C	.38	.17	.51	.22	C	3.91	14.09
3	0	143	36.7		-.31	-.13	-.40	-.17		2.62	12.74
4	0	77	19.7		-.24	-.12	-.34	-.17		2.55	12.55
OTHER	0	5	1.3		-.01	-.01	-.04	-.03		3.00	13.00
TOTAL		390									

1 LERTAP 2.0 — SUMMARY ITEM STATISTICS
PAGE 20

TEST NO 1 TTW.SH — SUBTEST 3 ID3 SUBTEST

ITEM NUMBER 21 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	B-ST	B-TT		ST	TT
1	0	80	20.5		-.23	-.03	-.32	-.05		2.58	13.06
C 2	1	91	23.3	C	.20	.01	.28	.02	C	3.58	13.32
3	0	128	32.8		-.22	-.11	-.29	-.14		2.74	12.80
C 4	1	88	22.6	C	.26	.16	.37	.22	C	3.73	14.11
OTHER	0	3	.8		-.01	-.07	-.04	-.28		3.00	11.00
TOTAL		390									

ITEM NUMBER 24 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	B-ST	B-TT		ST	TT
1	0	65	16.7		-.32	-.21	-.47	-.31		2.26	11.89
C 2	1	172	44.1	C	.23	.14	.29	.18	C	3.45	13.73
3	0	60	15.4		-.15	-.12	-.23	-.18		2.68	12.42
C 4	1	86	22.1	C	.16	.14	.22	.20	C	3.50	14.06
OTHER	0	7	1.8		-.06	-.08	-.18	-.23		2.57	11.57
TOTAL		390									

1 LERTAP 2.0 — SUBTEST STATISTICS
PAGE 21

TEST NO 1 TTW.SH — SUBTEST 3 ID3 SUBTEST

NUMBER OF INDIVIDUALS = 390.00 — NUMBER OF ITEMS = 6.00

MEAN = 3.13 — HIGHEST SCORE = 6.00

STANDARD DEVIATION = 1.23 — LOWEST SCORE = .00

SOURCE OF VARIANCE	D.F.	S.S.	M.S.
INDIVIDUALS	389.00	98.26	.25
ITEMS	5.00	13.87	2.77
RESIDUAL	1945.00	471.80	.24
TOTAL	2339.00	583.93	.25

HOYT ESTIMATE OF RELIABILITY = .04

STANDARD ERROR OF MEASUREMENT = 1.10

1 LERTAP 2.0 — SUMMARY ITEM STATISTICS

PAGE 22

TEST NO 1 TTW.SH SUBTEST 4 IIB3 SUBTEST

ITEM NUMBER 4 COEFFICIENTS OF CORRELATION MEANS

OPTION	WT	N	P		PB-ST	PB-TT	B-ST	B-TT		ST	TT
1	0	101	25.9		-.37	-.12	-.50	-.16		.39	12.67
2	0	70	17.9		-.26	-.22	-.38	-.33		.44	11.84
3	0	65	16.7		-.18	-.13	-.27	-.19		.55	12.40
C 4	1	151	38.7	C	.67	.38	.85	.48	C	1.54	14.66
OTHER	0	3	.8		.05	.01	.21	.05		1.33	13.67
TOTAL		390									

ITEM NUMBER 9 COEFFICIENTS OF CORRELATION MEANS

OPTION	WT	N	P		PB-ST	PB-TT	B-ST	B-TT		ST	TT
1	0	75	19.2		-.11	-.10	-.16	-.15		.69	12.64
C 2	1	53	13.6	C	.46	.02	.73	.03	C	1.79	13.40
3	0	187	47.9		-.19	.07	-.24	.09		.72	13.48
4	0	72	18.5		-.05	.00	-.08	.00		.79	13.24
OTHER	0	3	.8		.01	-.01	.06	-.03		1.00	13.00
TOTAL		390									

ITEM NUMBER 26 COEFFICIENTS OF CORRELATION MEANS

OPTION	WT	N	P		PB-ST	PB-TT	B-ST	B-TT		ST	TT
1	0	59	15.1		-.20	-.10	-.30	-.15		.51	12.56
2	0	135	34.6		-.33	-.05	-.43	-.07		.52	13.04
C 3	1	138	35.4	C	.63	.24	.81	.31	C	1.54	14.24
4	0	55	14.1		-.20	-.16	-.31	-.26		.49	12.05
OTHER	0	3	.8		-.02	.00	-.10	.01		.67	13.33
TOTAL		390									

1 LERTAP 2.0 SUBTEST STATISTICS
PAGE 23

TEST NO 1 TTW.SH SUBTEST 4 IIB3 SUBTEST

NUMBER OF INDIVIDUALS = 390.00 NUMBER OF ITEMS = 3.00

MEAN = .88 HIGHEST SCORE = 3.00

STANDARD DEVIATION = .79 LOWEST SCORE = .00

SOURCE OF VARIANCE	D.F.	S.S.	M.S.
INDIVIDUALS	389.00	80.03	.21
ITEMS	2.00	14.53	7.26
RESIDUAL	778.00	147.47	.19
TOTAL	1169.00	242.03	.21

HOYT ESTIMATE OF RELIABILITY = .08

STANDARD ERROR OF MEASUREMENT = .62

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 24

TEST NO 1 TTW.SH SUBTEST 5 IIB4 SUBTEST

ITEM NUMBER 2 COEFFICIENTS OF CORRELATION MEANS

OPTION	WT	N	P		PB-ST	PB-TT	B-ST	B-TT		ST	TT
1	0	28	7.2		-.18	-.08	-.33	-.16		1.46	12.36
C 2	1	251	64.4	C	.53	.29	.68	.37	C	2.70	13.89
3	0	44	11.3		-.30	-.19	-.49	-.32		1.23	11.64
4	0	67	17.2		-.30	-.14	-.45	-.21		1.43	12.31
TOTAL		390									

ITEM NUMBER 12					COEFFICIENTS OF CORRELATION					MEANS	
OPTION	WT	N	P		PB-ST	PB-TT	B-ST	B-TT		ST	TT
C 1	1	196	50.3	C	.53	.31	.66	.39	C	2.85	14.18
2	0	80	20.5		-.19	-.13	-.26	-.18		1.79	12.52
3	0	41	10.5		-.27	-.17	-.45	-.29		1.29	11.76
4	0	71	18.2		-.27	-.13	-.39	-.19		1.55	12.42
OTHER	0	2	.5		-.04	-.03	-.21	-.15		1.50	12.00
TOTAL		390									

ITEM NUMBER 15					COEFFICIENTS OF CORRELATION					MEANS	
OPTION	WT	N	P		PB-ST	PB-TT	B-ST	B-TT		ST	TT
1	0	114	29.2		-.14	-.02	-.19	-.03		1.96	13.17
2	0	45	11.5		-.17	-.18	-.28	-.30		1.67	11.78
C 3	1	114	29.2	C	.50	.29	.67	.38	C	3.17	14.60
4	0	111	28.5		-.23	-.13	-.30	-.18		1.79	12.63
OTHER	0	6	1.5		-.06	-.05	-.19	-.15		1.67	12.17
TOTAL		390									

ITEM NUMBER 20					COEFFICIENTS OF CORRELATION					MEANS	
OPTION	WT	N	P		PB-ST	PB-TT	B-ST	B-TT		ST	TT
C 1	1	165	42.3	C	.51	.32	.64	.41	C	2.94	14.38
2	0	73	18.7		-.22	-.20	-.32	-.29		1.67	12.01
3	0	86	22.1		-.30	-.16	-.41	-.23		1.56	12.34
4	0	63	16.2		-.11	-.03	-.17	-.04		1.92	13.08
OTHER	0	3	.8		-.02	-.04	-.07	-.15		2.00	12.00
TOTAL		390									

ITEM NUMBER 25					COEFFICIENTS OF CORRELATION					MEANS	
OPTION	WT	N	P		PB-ST	PB-TT	B-ST	B-TT		ST	TT
C 1	1	142	16.4	C	.42	.25	.54	.32	C	2.89	14.23
2	0	51	13.1		-.16	-.11	-.25	-.18		1.75	12.39
3	0	88	22.6		-.12	-.02	-.16	-.03		1.97	13.15
4	0	102	26.2		-.21	-.14	-.28	-.19		1.80	12.56
OTHER	0	7	1.8		-.06	-.09	-.17	-.27		1.71	11.29
TOTAL		390									

NUMBER OF INDIVIDUALS = 390.00 NUMBER OF ITEMS = 5.00

MEAN = 2.23 HIGHEST SCORE = 5.00

STANDARD DEVIATION = 1.20 LOWEST SCORE = .00

SOURCE OF VARIANCE	D.F.	S.S.	M.S.
INDIVIDUALS	389.00	112.03	.29
ITEMS	4.00	28.51	7.13
RESIDUAL	1556.00	341.09	.22
TOTAL	1949.00	481.63	.25

HOYT ESTIMATE OF RELIABILITY = .24

STANDARD ERROR OF MEASUREMENT = .94

NUMBER OF INDIVIDUALS = 390.00 NUMBER OF ITEMS = 26.00

MEAN = 13.26 HIGHEST SCORE = 22.00

STANDARD DEVIATION = 2.97 LOWEST SCORE = 4.00

SOURCE OF VARIANCE	D.F.	S.S.	M.S.
INDIVIDUALS	389.00	132.24	.34
ITEMS	25.00	334.05	13.36
RESIDUAL	9725.00	2067.71	.21
TOTAL	10139.00	2534.01	.25

HOYT ESTIMATE OF RELIABILITY = .37

STANDARD ERROR OF MEASUREMENT = 2.31

NO. SUBTESTS WITH NON-ZERO WT = 5.00

CRONBACHS ALPHA FOR COMPOSITE = .35

1 LERTAP 2.0 HISTOGRAM
PAGE 28

TEST NO 1 TTW.SH SUBTEST 1 ID1 SUBTEST

NUMBER OF OBSERVATIONS = 390 MEAN = 3.02 S.D. = 1.15 LOWEST SCORE = .00, HIGHEST SCORE= 6.00

Z	LB	UB	F	P	CF	CP	EACH * REPRESENTS 3 OBSERVATIONS
-16.47	-16.50	-15.50	0	.00	0	.00	
-15.60	-15.50	-14.50	0	.00	0	.00	
-14.74	-14.50	-13.50	0	.00	0	.00	
-13.87	-13.50	-12.50	0	.00	0	.00	
-13.00	-12.50	-11.50	0	.00	0	.00	
-12.14	-11.50	-10.50	0	.00	0	.00	
-11.27	-10.50	-9.50	0	.00	0	.00	
-10.41	-9.50	-8.50	0	.00	0	.00	
-9.54	-8.50	-7.50	0	.00	0	.00	
-8.67	-7.50	-6.50	0	.00	0	.00	
-7.81	-6.50	-5.50	0	.00	0	.00	
-6.94	-5.50	-4.50	0	.00	0	.00	
-6.08	-4.50	-3.50	0	.00	0	.00	
-5.21	-3.50	-2.50	0	.00	0	.00	
-4.34	-2.50	-1.50	0	.00	0	.00	
-3.48	-1.50	-.50	0	.00	0	.00	
-2.61	-.50	.50	5	1.28	5	1.28	**
-1.75	.50	1.50	31	7.95	36	9.23	**********
-.88	1.50	2.50	84	21.54	120	30.77	**********
-.02	2.50	3.50	144	36.92	264	67.69	**********
.85	3.50	4.50	91	23.33	355	91.03	**********
1.72	4.50	5.50	28	7.18	383	98.21	*********
2.58	5.50	6.50	7	1.79	390	100.00	**
3.45	6.50	7.50	0	.00	390	100.00	
4.31	7.50	8.50	0	.00	390	100.00	
5.18	8.50	9.50	0	.00	390	100.00	
6.05	9.50	10.50	0	.00	390	100.00	
6.91	10.50	11.50	0	.00	390	100.00	
7.78	11.50	12.50	0	.00	390	100.00	
8.64	12.50	13.50	0	.00	390	100.00	
9.51	13.50	14.50	0	.00	390	100.00	
10.37	14.50	15.50	0	.00	390	100.00	
11.24	15.50	16.50	0	.00	390	100.00	
12.11	16.50	17.50	0	.00	390	100.00	
12.97	17.50	18.50	0	.00	390	100.00	
13.84	18.50	19.50	0	.00	390	100.00	
14.70	19.50	20.50	0	.00	390	100.00	
15.57	20.50	21.50	0	.00	390	100.00	
16.44	21.50	22.50	0	.00	390	100.00	
17.30	22.50	23.50	0	.00	390	100.00	

1 LERTAP 2.0 HISTOGRAM
PAGE 29

TEST NO 1 TTW.SH SUBTEST 2 ID2 SUBTEST

NUMBER OF OBSERVATIONS = 390 MEAN = 4.01 S.D. = 1.21 LOWEST SCORE = .00, HIGHEST SCORE= 6.00

Z	LB	UB	F	P	CF	CP	EACH * REPRESENTS 2 OBSERVATIONS
-16.49	-16.50	-15.50	0	.00	0	.00	
-15.66	-15.50	-14.50	0	.00	0	.00	
-14.84	-14.50	-13.50	0	.00	0	.00	
-14.01	-13.50	-12.50	0	.00	0	.00	
-13.19	-12.50	-11.50	0	.00	0	.00	
-12.37	-11.50	-10.50	0	.00	0	.00	
-11.54	-10.50	-9.50	0	.00	0	.00	
-10.72	-9.50	-8.50	0	.00	0	.00	
-9.89	-8.50	-7.50	0	.00	0	.00	
-9.07	-7.50	-6.50	0	.00	0	.00	
-8.25	-6.50	-5.50	0	.00	0	.00	
-7.42	-5.50	-4.50	0	.00	0	.00	
-6.60	-4.50	-3.50	0	.00	0	.00	
-5.77	-3.50	-2.50	0	.00	0	.00	
-4.95	-2.50	-1.50	0	.00	0	.00	
-4.13	-1.50	-.50	0	.00	0	.00	
-3.30	-.50	.50	1	.26	1	.26	*
-2.48	.50	1.50	6	1.54	7	1.79	***
-1.65	1.50	2.50	41	10.51	48	12.31	**********
-.83	2.50	3.50	78	20.00	126	32.31	**********
-.01	3.50	4.50	117	30.00	243	62.31	**********
.82	4.50	5.50	109	27.95	352	90.26	**********
1.64	5.50	6.50	38	9.74	390	100.00	**********
2.47	6.50	7.50	0	.00	390	100.00	
3.29	7.50	8.50	0	.00	390	100.00	
4.11	8.50	9.50	0	.00	390	100.00	
4.94	9.50	10.50	0	.00	390	100.00	
5.76	10.50	11.50	0	.00	390	100.00	
6.59	11.50	12.50	0	.00	390	100.00	
7.41	12.50	13.50	0	.00	390	100.00	
8.23	13.50	14.50	0	.00	390	100.00	
9.06	14.50	15.50	0	.00	390	100.00	
9.88	15.50	16.50	0	.00	390	100.00	
10.71	16.50	17.50	0	.00	390	100.00	
11.53	17.50	18.50	0	.00	390	100.00	
12.35	18.50	19.50	0	.00	390	100.00	
13.18	19.50	20.50	0	.00	390	100.00	
14.00	20.50	21.50	0	.00	390	100.00	
14.82	21.50	22.50	0	.00	390	100.00	
15.65	22.50	23.50	0	.00	390	100.00	

1 LERTAP 2.0 HISTOGRAM
PAGE 30

TEST NO 1 TTW.SH SUBTEST 3 ID3 SUBTEST

NUMBER OF OBSERVATIONS = 390 MEAN = 3.13 S.D. = 1.23 LOWEST SCORE = .00, HIGHEST SCORE= 6.00

Z	LB	UB	F	P	CF	CP	EACH * REPRESENTS 2 OBSERVATIONS
-15.54	-16.50	-15.50	0	.00	0	.00	
-14.72	-15.50	-14.50	0	.00	0	.00	
-13.91	-14.50	-13.50	0	.00	0	.00	
-13.10	-13.50	-12.50	0	.00	0	.00	
-12.29	-12.50	-11.50	0	.00	0	.00	
-11.48	-11.50	-10.50	0	.00	0	.00	
-10.66	-10.50	-9.50	0	.00	0	.00	
-9.85	-9.50	-8.50	0	.00	0	.00	
-9.04	-8.50	-7.50	0	.00	0	.00	
-8.23	-7.50	-6.50	0	.00	0	.00	
-7.41	-6.50	-5.50	0	.00	0	.00	
-6.60	-5.50	-4.50	0	.00	0	.00	
-5.79	-4.50	-3.50	0	.00	0	.00	
-4.98	-3.50	-2.50	0	.00	0	.00	
-4.17	-2.50	-1.50	0	.00	0	.00	
-3.35	-1.50	-.50	0	.00	0	.00	
-2.54	-.50	.50	5	1.28	5	1.28	***
-1.73	.50	1.50	33	8.46	38	9.74	*****************
-.92	1.50	2.50	79	20.26	117	30.00	**
-.10	2.50	3.50	121	31.03	238	61.03	**
.71	3.50	4.50	101	25.90	339	86.92	***
1.52	4.50	5.50	44	11.28	383	98.21	**********************
2.33	5.50	6.50	7	1.79	390	100.00	****
3.14	6.50	7.50	0	.00	390	100.00	
3.96	7.50	8.50	0	.00	390	100.00	
4.77	8.50	9.50	0	.00	390	100.00	
5.58	9.50	10.50	0	.00	390	100.00	
6.39	10.50	11.50	0	.00	390	100.00	
7.21	11.50	12.50	0	.00	390	100.00	
8.02	12.50	13.50	0	.00	390	100.00	
8.83	13.50	14.50	0	.00	390	100.00	
9.64	14.50	15.50	0	.00	390	100.00	
10.46	15.50	16.50	0	.00	390	100.00	
11.27	16.50	17.50	0	.00	390	100.00	
12.08	17.50	18.50	0	.00	390	100.00	
12.89	18.50	19.50	0	.00	390	100.00	
13.70	19.50	20.50	0	.00	390	100.00	
14.52	20.50	21.50	0	.00	390	100.00	
15.33	21.50	22.50	0	.00	390	100.00	
16.14	22.50	23.50	0	.00	390	100.00	

1 LERTAP 2.0 HISTOGRAM
PAGE 31

TEST NO 1 TTW.SH SUBTEST 4 IIB3 SUBTEST

NUMBER OF OBSERVATIONS = 390 MEAN = .88 S.D. = .79 LOWEST SCORE = .00, HIGHEST SCORE= 3.00

Z	LB	UB	F	P	CF	CP	EACH * REPRESENTS 3 OBSERVATIONS
-24.03	-18.50	-17.50	0	.00	0	.00	
-22.76	-17.50	-16.50	0	.00	0	.00	
-21.48	-16.50	-15.50	0	.00	0	.00	
-20.21	-15.50	-14.50	0	.00	0	.00	
-18.94	-14.50	-13.50	0	.00	0	.00	
-17.66	-13.50	-12.50	0	.00	0	.00	
-16.39	-12.50	-11.50	0	.00	0	.00	
-15.12	-11.50	-10.50	0	.00	0	.00	
-13.84	-10.50	-9.50	0	.00	0	.00	
-12.57	-9.50	-8.50	0	.00	0	.00	
-11.30	-8.50	-7.50	0	.00	0	.00	
-10.03	-7.50	-6.50	0	.00	0	.00	
-8.75	-6.50	-5.50	0	.00	0	.00	
-7.48	-5.50	-4.50	0	.00	0	.00	
-6.21	-4.50	-3.50	0	.00	0	.00	
-4.93	-3.50	-2.50	0	.00	0	.00	
-3.66	-2.50	-1.50	0	.00	0	.00	
-2.39	-1.50	-.50	0	.00	0	.00	
-1.12	-.50	.50	139	35.64	139	35.64	**********
.16	.50	1.50	168	43.08	307	78.72	**********
1.43	1.50	2.50	75	19.23	382	97.95	**********
2.70	2.50	3.50	8	2.05	390	100.00	***
3.98	3.50	4.50	0	.00	390	100.00	
5.25	4.50	5.50	0	.00	390	100.00	
6.52	5.50	6.50	0	.00	390	100.00	
7.79	6.50	7.50	0	.00	390	100.00	
9.07	7.50	8.50	0	.00	390	100.00	
10.34	8.50	9.50	0	.00	390	100.00	
11.61	9.50	10.50	0	.00	390	100.00	
12.89	10.50	11.50	0	.00	390	100.00	
14.16	11.50	12.50	0	.00	390	100.00	
15.43	12.50	13.50	0	.00	390	100.00	
16.70	13.50	14.50	0	.00	390	100.00	
17.98	14.50	15.50	0	.00	390	100.00	
19.25	15.50	16.50	0	.00	390	100.00	
20.52	16.50	17.50	0	.00	390	100.00	
21.80	17.50	18.50	0	.00	390	100.00	
23.07	18.50	19.50	0	.00	390	100.00	
24.34	19.50	20.50	0	.00	390	100.00	
25.61	20.50	21.50	0	.00	390	100.00	

1 LERTAP 2.0 HISTOGRAM
PAGE 32

TEST NO 1 TTW.SH SUBTEST 5 IIB4 SUBTEST

NUMBER OF OBSERVATIONS = 390 MEAN = 2.23 S.D. = 1.20 LOWEST SCORE = .00, HIGHEST SCORE= 5.00

Z	LB	UB	F	P	CF	CP	EACH * REPRESENTS 2 OBSERVATIONS
-16.02	-17.50	-16.50	0	.00	0	.00	
-15.19	-16.50	-15.50	0	.00	0	.00	
-14.35	-15.50	-14.50	0	.00	0	.00	
-13.52	-14.50	-13.50	0	.00	0	.00	
-12.69	-13.50	-12.50	0	.00	0	.00	
-11.85	-12.50	-11.50	0	.00	0	.00	
-11.02	-11.50	-10.50	0	.00	0	.00	
-10.19	-10.50	-9.50	0	.00	0	.00	
-9.35	-9.50	-8.50	0	.00	0	.00	
-8.52	-8.50	-7.50	0	.00	0	.00	
-7.69	-7.50	-6.50	0	.00	0	.00	
-6.85	-6.50	-5.50	0	.00	0	.00	
-6.02	-5.50	-4.50	0	.00	0	.00	
-5.19	-4.50	-3.50	0	.00	0	.00	
-4.35	-3.50	-2.50	0	.00	0	.00	
-3.52	-2.50	-1.50	0	.00	0	.00	
-2.69	-1.50	-.50	0	.00	0	.00	
-1.85	-.50	.50	26	6.67	26	6.67	*************
-1.02	.50	1.50	87	22.31	113	28.97	**********
-.19	1.50	2.50	117	30.00	230	58.97	**********
.65	2.50	3.50	107	27.44	337	86.41	**********
1.48	3.50	4.50	39	10.00	376	96.41	**********
2.31	4.50	5.50	14	3.59	390	100.00	*******
3.15	5.50	6.50	0	.00	390	100.00	
3.98	6.50	7.50	0	.00	390	100.00	
4.81	7.50	8.50	0	.00	390	100.00	
5.65	8.50	9.50	0	.00	390	100.00	
6.48	9.50	10.50	0	.00	390	100.00	
7.31	10.50	11.50	0	.00	390	100.00	
8.15	11.50	12.50	0	.00	390	100.00	
8.98	12.50	13.50	0	.00	390	100.00	
9.81	13.50	14.50	0	.00	390	100.00	
10.65	14.50	15.50	0	.00	390	100.00	
11.48	15.50	16.50	0	.00	390	100.00	
12.31	16.50	17.50	0	.00	390	100.00	
13.15	17.50	18.50	0	.00	390	100.00	
13.98	18.50	19.50	0	.00	390	100.00	
14.81	19.50	20.50	0	.00	390	100.00	
15.65	20.50	21.50	0	.00	390	100.00	
16.48	21.50	22.50	0	.00	390	100.00	

1 LERTAP 2.0 HISTOGRAM
PAGE 33

TEST NO 1 TTW.SH TOTAL TEST

NUMBER OF OBSERVATIONS = 390 MEAN = 13.26 S.D. = 2.97 LOWEST SCORE = 4.00, HIGHEST SCORE= 22.00

Z	LB	UB	F	P	CF	CP	EACH * REPRESENTS 1 OBSERVATIONS
-6.48	-6.50	-5.50	0	.00	0	.00	
-6.14	-5.50	-4.50	0	.00	0	.00	
-5.80	-4.50	-3.50	0	.00	0	.00	
-5.47	-3.50	-2.50	0	.00	0	.00	
-5.13	-2.50	-1.50	0	.00	0	.00	
-4.80	-1.50	-.50	0	.00	0	.00	
-4.46	-.50	.50	0	.00	0	.00	
-4.12	.50	1.50	0	.00	0	.00	
-3.79	1.50	2.50	0	.00	0	.00	
-3.45	2.50	3.50	0	.00	0	.00	
-3.11	3.50	4.50	2	.51	2	.51	**
-2.78	4.50	5.50	0	.00	2	.51	
-2.44	5.50	6.50	2	.51	4	1.03	**
-2.10	6.50	7.50	5	1.28	9	2.31	*****
-1.77	7.50	8.50	7	1.79	16	4.10	*******
-1.43	8.50	9.50	25	6.41	41	10.51	*************************
-1.10	9.50	10.50	34	8.72	75	19.23	**********************************
-.76	10.50	11.50	35	8.97	110	28.21	***********************************
-.42	11.50	12.50	47	12.05	157	40.26	**
-.09	12.50	13.50	50	12.82	207	53.08	**
.25	13.50	14.50	45	11.54	252	64.62	**
.59	14.50	15.50	48	12.31	300	76.92	**
.92	15.50	16.50	35	8.97	335	85.90	***********************************
1.26	16.50	17.50	27	6.92	362	92.82	***************************
1.60	17.50	18.50	16	4.10	378	96.92	****************
1.93	18.50	19.50	6	1.54	384	98.46	******
2.27	19.50	20.50	3	.77	387	99.23	***
2.60	20.50	21.50	2	.51	389	99.74	**
2.94	21.50	22.50	1	.26	390	100.00	*
3.28	22.50	23.50	0	.00	390	100.00	
3.61	23.50	24.50	0	.00	390	100.00	
3.95	24.50	25.50	0	.00	390	100.00	
4.29	25.50	26.50	0	.00	390	100.00	
4.62	26.50	27.50	0	.00	390	100.00	
4.96	27.50	28.50	0	.00	390	100.00	
5.30	28.50	29.50	0	.00	390	100.00	
5.63	29.50	30.50	0	.00	390	100.00	
5.97	30.50	31.50	0	.00	390	100.00	
6.30	31.50	32.50	0	.00	390	100.00	
6.64	32.50	33.50	0	.00	390	100.00	

1 LERTAP 2.0 CORRELATIONS
PAGE 34

TEST NO 1 TTW.SH

VARIABLE	TYPE	NAME
1	SUBTEST 1	ID1 SUBTEST
2	SUBTEST 2	ID2 SUBTEST
3	SUBTEST 3	ID3 SUBTEST
4	SUBTEST 4	IIB3 SUBTEST
5	SUBTEST 5	IIB4 SUBTEST
6	TOTAL TEST	

1 LERTAP 2.0 CORRELATIONS
PAGE 35

VAR	1	2	3	4	5	6
1	1.000	.192	.054	.022	.171	.565
2	.192	1.000	.025	.117	.205	.607
3	.054	.025	1.000	.054	.008	.463
4	.022	.117	.054	1.000	.120	.391
5	.171	.205	.008	.120	1.000	.589
6	.565	.607	.463	.391	.589	1.000

1

APPENDIX E

RESULTS OF ITEM ANALYSIS FOR THE TOEFL

TEST NO 1 TOEFL — SUBTEST 1 SECTION 1: LISTENING

ITEM NUMBER 1

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		MEANS ST	MEANS TT	MEANS EC
1	0	1	.3		.02	.03	.06	.10	.20	.38		36.00	116.00	17.00
2	0	2	.5		-.14	-.13	.04	-.67	-.63	.19		18.00	72.00	15.00
3	0	105	26.9		-.42	-.36	-.14	-.56	-.49	-.18		27.91	94.18	12.45
C 4	1	282	72.3	C	.44	.38	.12	.58	.50	.16	C	35.65	109.28	13.43
TOTAL		390												

ITEM NUMBER 2

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		MEANS ST	MEANS TT	MEANS EC
1	0	15	3.8		-.18	-.17	-.04	-.41	-.39	-.10		26.40	89.60	12.47
2	0	72	18.5		-.20	-.19	-.19	-.30	-.28	-.27		30.01	97.76	11.90
3	0	7	1.8		-.13	-.13	-.01	-.39	-.39	-.04		25.71	87.86	12.86
C 4	1	296	75.9	C	.31	.29	.19	.42	.40	.26	C	34.86	108.00	13.53
TOTAL		390												

ITEM NUMBER 3

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		MEANS ST	MEANS TT	MEANS EC
1	0	56	14.4		-.13	-.10	.01	-.20	-.16	.01		30.98	100.61	13.23
C 2	1	266	68.2	C	.32	.30	.05	.42	.39	.07	C	35.24	108.74	13.29
3	0	55	14.1		-.25	-.23	-.05	-.38	-.36	-.08		28.56	94.73	12.76
4	0	13	3.3		-.10	-.13	-.05	-.25	-.33	-.11		29.00	91.92	12.38
TOTAL		390												

ITEM NUMBER 4

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		MEANS ST	MEANS TT	MEANS EC
1	0	2	.5		-.12	-.13	-.07	-.58	-.61	-.34		20.00	73.00	10.00
2	0	15	3.8		-.24	-.23	-.10	-.56	-.54	-.23		23.73	83.93	11.53
C 3	1	356	91.3	C	.40	.39	.15	.66	.65	.24	C	34.47	107.21	13.33
4	0	17	4.4		-.28	-.27	-.08	-.62	-.60	-.18		22.94	82.00	11.94
TOTAL		390												

ITEM NUMBER 5

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		MEANS ST	MEANS TT	MEANS EC
1	0	32	8.2		-.19	-.17	-.19	-.35	-.30	-.34		28.31	95.03	11.16
2	0	19	4.9		-.27	-.26	-.08	-.58	-.55	-.17		23.84	84.26	12.00
C 3	1	331	84.9	C	.37	.33	.18	.55	.49	.26	C	34.73	107.56	13.42
4	0	8	2.1		-.15	-.12	.04	-.42	-.34	.10		25.25	90.25	14.00
TOTAL		390												

TEST NO 1 TOEFL — SUBTEST 1 SECTION 1: LISTENING

ITEM NUMBER 6

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		MEANS ST	MEANS TT	MEANS EC
C 1	1	263	67.4	C	.44	.43	.17	.57	.56	.23	C	35.93	110.42	13.57
2	0	80	20.5		-.29	-.27	-.12	-.41	-.39	-.16		28.91	95.26	12.44
3	0	21	5.4		-.17	-.18	-.04	-.35	-.36	-.09		27.81	91.71	12.57
4	0	25	6.4		-.21	-.20	-.09	-.41	-.39	-.18		27.00	91.08	12.00
OTHER	0	1	.3		.00	-.01	-.03	.02	-.07	-.22		34.00	101.00	11.00
TOTAL		390												

ITEM NUMBER 7

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		MEANS ST	MEANS TT	MEANS EC
C 1	1	318	81.5	C	.44	.41	.19	.64	.60	.28	C	35.17	108.61	13.48
2	0	9	2.3		-.09	-.04	-.03	-.24	-.12	-.10		29.00	99.89	12.44
3	0	44	11.3		-.33	-.32	-.13	-.54	-.53	-.22		26.14	88.75	11.95
4	0	19	4.9		-.26	-.24	-.13	-.56	-.52	-.27		24.21	85.47	11.37
TOTAL		390												

ITEM NUMBER 8

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
C 1	1	336	86.2	C	.32	.32	.12	.48	.48	.18	C	34.50	107.33	13.33
2	0	1	.3		-.08	-.15	-.06	-.50	-.98	-.41		21.00	50.00	9.00
3	0	12	3.1		-.17	-.19	-.14	-.43	-.47	-.34		25.67	85.75	10.67
4	0	41	10.5		-.25	-.22	-.04	-.42	-.38	-.07		27.66	93.22	12.78
TOTAL		390												

ITEM NUMBER 9

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	120	30.8		-.14	-.12	-.01	-.19	-.16	-.01		31.74	101.71	13.13
2	0	24	6.2		-.21	-.15	-.07	-.41	-.30	-.14		26.92	94.29	12.25
3	0	20	5.1		-.15	-.15	-.01	-.32	-.31	-.02		28.20	93.40	13.05
C 4	1	226	57.9	C	.30	.26	.05	.38	.32	.06	C	35.56	108.98	13.31
TOTAL		390												

ITEM NUMBER 10

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	6	1.5		-.17	-.20	-.09	-.53	-.62	-.29		22.83	76.67	10.83
C 2	1	331	84.9	C	.46	.43	.14	.68	.65	.20	C	35.03	108.34	13.37
3	0	18	4.6		-.29	-.28	-.02	-.63	-.60	-.03		22.89	82.22	12.94
4	0	35	9.0		-.29	-.25	-.12	-.51	-.45	-.21		26.03	90.43	11.94
TOTAL		390												

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 12

TEST NO 1 TOEFL SUBTEST 1 SECTION 1: LISTENING

ITEM NUMBER 11

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	31	7.9		-.20	-.16	.02	-.36	-.29	.03		28.13	95.23	13.39
2	0	38	9.7		-.34	-.27	-.17	-.58	-.47	-.29		25.18	90.08	11.50
3	0	28	7.2		-.23	-.21	-.04	-.44	-.40	-.07		26.68	91.25	12.75
C 4	1	293	75.1	C	.49	.41	.13	.67	.56	.17	C	35.77	109.33	13.42
TOTAL		390												

ITEM NUMBER 12

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	15	3.8		-.23	-.24	-.14	-.53	-.56	-.32		24.20	82.93	10.93
C 2	1	348	89.2	C	.43	.42	.18	.69	.67	.28	C	34.69	107.68	13.38
3	0	13	3.3		-.24	-.18	-.08	-.58	-.43	-.20		23.15	87.85	11.69
4	0	14	3.6		-.25	-.28	-.07	-.60	-.66	-.16		22.93	78.93	12.00
TOTAL		390												

ITEM NUMBER 13

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	11	2.8		-.13	-.13	-.10	-.34	-.34	-.26		27.18	90.82	11.27
2	0	28	7.2		-.18	-.20	-.10	-.34	-.38	-.20		28.25	92.00	11.96
3	0	10	2.6		-.09	-.06	.02	-.23	-.15	.06		29.20	98.80	13.60
C 4	1	341	87.4	C	.25	.25	.12	.39	.39	.19	C	34.23	106.75	13.33
TOTAL		390												

ITEM NUMBER 14

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	22	5.6		-.17	-.18	.03	-.35	-.37	.06		27.82	91.45	13.55
2	0	3	.8		-.06	-.04	-.09	-.26	-.18	-.35		27.67	96.00	10.00
3	0	15	3.8		-.24	-.20	-.04	-.56	-.46	-.08		23.67	87.13	12.60
C 4	1	350	89.7	C	.30	.28	.03	.49	.45	.04	C	34.30	106.74	13.21
TOTAL		390												

ITEM NUMBER 15

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	64	16.4		-.21	-.16	-.07	-.31	-.24	-.10		29.73	98.53	12.69
2	0	53	13.6		-.14	-.17	-.05	-.22	-.27	-.08		30.64	97.26	12.75
C 3	1	237	60.8	C	.37	.35	.12	.48	.44	.15	C	35.90	110.06	13.49
4	0	36	9.2		-.20	-.18	-.05	-.35	-.31	-.09		28.36	95.00	12.67
TOTAL		390												

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 13

TEST NO 1 TOEFL SUBTEST 1 SECTION 1: LISTENING

ITEM NUMBER 16

				COEFFICIENTS OF CORRELATION						MEANS		
OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
C 1	1	140	35.9	C .29	.23	.04	.38	.30	.05	C 36.65	110.66	13.36
2	0	119	30.5	-.10	-.07	-.05	-.13	-.10	-.07	32.32	103.05	12.92
3	0	5	1.3	-.09	-.07	.01	-.31	-.25	.03	27.00	93.40	13.40
4	0	125	32.1	-.19	-.15	.01	-.25	-.20	.02	31.26	101.04	13.25
OTHER	0	1	.3	.02	.02	-.06	.14	.14	-.41	37.00	113.00	9.00
TOTAL		390										

ITEM NUMBER 17

				COEFFICIENTS OF CORRELATION						MEANS		
OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	12	3.1	-.25	-.24	-.05	-.61	-.58	-.11	22.33	81.17	12.33
2	0	25	6.4	-.17	-.17	-.12	-.34	-.33	-.24	28.12	93.52	11.64
C 3	1	311	79.7	C .33	.29	.09	.47	.41	.12	C 34.83	107.71	13.32
4	0	42	10.8	-.16	-.12	.01	-.26	-.20	.02	29.83	98.93	13.29
TOTAL		390										

ITEM NUMBER 18

				COEFFICIENTS OF CORRELATION						MEANS		
OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
C 1	1	293	75.1	C .49	.46	.18	.66	.62	.24	C 35.74	109.83	13.51
2	0	13	3.3	-.17	-.18	-.09	-.41	-.44	-.22	26.15	87.31	11.62
3	0	27	6.9	-.27	-.26	-.15	-.52	-.50	-.28	25.41	87.63	11.41
4	0	57	14.6	-.31	-.28	-.06	-.49	-.43	-.10	27.35	92.70	12.68
TOTAL		390										

ITEM NUMBER 19

				COEFFICIENTS OF CORRELATION						MEANS		
OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
C 1	1	271	69.5	C .41	.36	.14	.54	.47	.18	C 35.66	109.37	13.48
2	0	37	9.5	-.15	-.15	-.09	-.26	-.27	-.16	29.78	96.46	12.27
3	0	73	18.7	-.33	-.27	-.07	-.48	-.39	-.11	27.96	94.85	12.68
4	0	9	2.3	-.11	-.11	-.06	-.30	-.29	-.17	27.78	92.44	11.89
TOTAL		390										

ITEM NUMBER 20

				COEFFICIENTS OF CORRELATION						MEANS		
OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	64	16.4	-.33	-.28	-.04	-.49	-.41	-.06	27.48	93.80	12.89
2	0	45	11.5	-.30	-.25	-.16	-.49	-.42	-.26	26.89	92.27	11.76
C 3	1	244	62.6	C .54	.47	.16	.69	.59	.21	C 36.86	111.55	13.59
4	0	37	9.5	-.16	-.14	-.05	-.28	-.25	-.09	29.51	97.08	12.68
TOTAL		390										

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 14

TEST NO 1 TOEFL SUBTEST 1 SECTION 1: LISTENING

ITEM NUMBER 21

				COEFFICIENTS OF CORRELATION						MEANS		
OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	37	9.5	-.17	-.16	-.05	-.29	-.27	-.08	29.32	96.32	12.70
C 2	1	236	60.5	C .42	.35	.08	.53	.44	.11	C 36.21	110.10	13.40
3	0	84	21.5	-.28	-.23	-.05	-.39	-.32	-.06	29.17	97.08	12.89
4	0	32	8.2	-.13	-.08	-.02	-.24	-.15	-.03	29.94	100.00	13.00
OTHER	0	1	.3	-.10	-.12	-.08	-.62	-.75	-.51	18.00	63.00	8.00
TOTAL		390										

ITEM NUMBER 22

				COEFFICIENTS OF CORRELATION						MEANS		
OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	47	12.1	-.01	-.01	-.05	-.02	-.01	-.09	33.23	104.79	12.70
2	0	32	8.2	-.17	-.15	-.05	-.31	-.27	-.09	28.91	96.09	12.66
C 3	1	254	65.1	C .20	.18	.09	.26	.23	.11	C 34.67	107.16	13.39
4	0	57	14.6	-.13	-.12	-.03	-.20	-.18	-.05	30.91	99.93	12.93
TOTAL		390										

ITEM NUMBER 23

				COEFFICIENTS OF CORRELATION						MEANS		
OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
C 1	1	224	57.4	C .47	.39	.10	.59	.49	.12	C 36.71	111.15	13.45
2	0	42	10.8	-.28	-.23	-.09	-.47	-.38	-.16	26.95	93.31	12.31
3	0	33	8.5	-.14	-.16	-.11	-.25	-.30	-.19	29.79	95.24	12.03
4	0	91	23.3	-.25	-.19	.03	-.34	-.26	.04	29.87	98.97	13.34
TOTAL		390										

ITEM NUMBER 24 — COEFFICIENTS OF CORRELATION — MEANS

	OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
	1	0	11	2.8		-.01	-.01	-.06	-.02	-.04	-.15		33.09	103.45	12.09
C	2	1	240	61.5	C	.39	.32	.17	.50	.41	.22	C	35.95	109.65	13.62
	3	0	116	29.7		-.32	-.26	-.13	-.42	-.34	-.17		29.52	97.91	12.53
	4	0	23	5.9		-.18	-.16	-.06	-.36	-.32	-.11		27.78	93.65	12.43
	TOTAL		390												

ITEM NUMBER 25 — COEFFICIENTS OF CORRELATION — MEANS

	OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
	1	0	88	22.6		-.02	.03	.00	-.03	.04	.01		33.17	106.08	13.20
	2	0	20	5.1		-.06	-.05	-.03	-.13	-.11	-.07		31.25	100.85	12.70
C	3	1	197	50.5	C	.21	.14	.12	.26	.18	.15	C	35.15	107.56	13.57
	4	0	85	21.8		-.20	-.17	-.13	-.28	-.24	-.18		30.45	99.11	12.36
	TOTAL		390												

TEST NO 1 TOEFL — SUBTEST 1 SECTION 1: LISTENING

ITEM NUMBER 26 — COEFFICIENTS OF CORRELATION — MEANS

	OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
	1	0	21	5.4		-.23	-.20	-.05	-.48	-.41	-.10		25.67	89.81	12.52
	2	0	62	15.9		-.28	-.25	-.12	-.42	-.37	-.18		28.37	94.74	12.31
C	3	1	238	61.0	C	.47	.44	.15	.60	.56	.19	C	36.52	111.38	13.56
	4	0	69	17.7		-.20	-.20	-.05	-.30	-.30	-.07		29.96	97.06	12.84
	TOTAL		390												

ITEM NUMBER 27 — COEFFICIENTS OF CORRELATION — MEANS

	OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
	1	0	15	3.8		-.06	-.06	-.09	-.13	-.15	-.20		31.13	99.27	11.73
C	2	1	244	62.6	C	.18	.20	.07	.23	.25	.09	C	34.62	107.82	13.36
	3	0	65	16.7		-.18	-.20	-.14	-.26	-.30	-.20		30.29	96.77	12.18
	4	0	66	16.9		-.03	-.02	.09	-.05	-.03	.13		32.91	104.23	13.80
	TOTAL		390												

ITEM NUMBER 28 — COEFFICIENTS OF CORRELATION — MEANS

	OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
	1	0	66	16.9		-.18	-.11	-.02	-.26	-.16	-.03		30.33	100.71	13.03
	2	0	77	19.7		-.05	-.06	.00	-.08	-.08	.00		32.58	103.00	13.17
	3	0	83	21.3		-.17	-.17	-.13	-.25	-.24	-.18		30.77	99.00	12.36
C	4	1	164	42.1	C	.32	.27	.13	.41	.34	.16	C	36.53	110.79	13.66
	TOTAL		390												

ITEM NUMBER 29 — COEFFICIENTS OF CORRELATION — MEANS

	OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
C	1	1	166	42.6	C	.28	.19	.04	.35	.23	.05	C	36.07	108.95	13.14
	2	0	50	12.8		-.07	-.03	.04	-.11	-.05	.07		32.06	103.64	13.54
	3	0	67	17.2		-.23	-.21	-.11	-.34	-.32	-.16		29.40	96.58	12.42
	4	0	107	27.4		-.06	.00	.01	-.08	-.01	.01		32.66	104.93	13.23
	TOTAL		390												

ITEM NUMBER 30 — COEFFICIENTS OF CORRELATION — MEANS

	OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
	1	0	108	27.7		.03	.04	.01	.04	.05	.02		33.87	106.12	13.24
	2	0	143	36.7		-.03	-.01	.00	-.04	-.01	.00		33.13	104.90	13.18
	3	0	40	10.3		-.10	-.13	-.04	-.18	-.22	-.08		31.00	98.00	12.75
C	4	1	99	25.4	C	.08	.06	.02	.11	.08	.03	C	34.56	106.91	13.28
	TOTAL		390												

TEST NO 1 TOEFL — SUBTEST 1 SECTION 1: LISTENING

ITEM NUMBER 31 — COEFFICIENTS OF CORRELATION — MEANS

	OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
	1	0	23	5.9		-.25	-.27	-.03	-.50	-.54	-.06		25.39	85.65	12.78
	2	0	11	2.8		-.21	-.20	-.14	-.54	-.51	-.35		23.55	83.73	10.55
	3	0	49	12.6		-.27	-.28	-.09	-.43	-.44	-.15		27.80	91.82	12.37
C	4	1	307	78.7	C	.45	.46	.15	.62	.64	.21	C	35.35	109.36	13.43
	TOTAL		390												

ITEM NUMBER 32

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
C 1	1	252	64.6	C	.48	.46	.19	.62	.59	.25	C	36.35	111.21	13.64
2	0	49	12.6		-.24	-.27	-.09	-.38	-.43	-.15		28.39	92.33	12.39
3	0	43	11.0		-.20	-.19	-.07	-.33	-.32	-.12		29.00	95.19	12.51
4	0	46	11.8		-.28	-.22	-.12	-.46	-.36	-.20		27.33	94.00	12.11
TOTAL		390												

ITEM NUMBER 33

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	21	5.4		-.08	-.10	-.09	-.17	-.20	-.19		30.67	97.52	11.90
2	0	52	13.3		-.28	-.24	-.06	-.43	-.37	-.10		27.83	94.10	12.67
3	0	89	22.8		-.24	-.22	-.10	-.34	-.30	-.13		29.90	97.84	12.61
C 4	1	227	58.2	C	.45	.40	.18	.56	.51	.22	C	36.52	111.22	13.67
OTHER	0	1	.3		-.12	-.10	-.10	-.74	-.63	-.61		15.00	70.00	7.00
TOTAL		390												

ITEM NUMBER 34

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
C 1	1	344	88.2	C	.41	.42	.18	.65	.65	.28	C	34.69	107.79	13.40
2	0	6	1.5		-.23	-.25	-.07	-.73	-.78	-.23		18.67	69.50	11.33
3	0	29	7.4		-.26	-.28	-.13	-.48	-.52	-.24		26.14	87.41	11.72
4	0	11	2.8		-.22	-.19	-.10	-.56	-.49	-.26		23.09	84.73	11.27
TOTAL		390												

ITEM NUMBER 35

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	26	6.7		-.15	-.14	-.09	-.29	-.28	-.17		28.96	95.27	12.12
C 2	1	262	67.2	C	.48	.41	.14	.62	.53	.18	C	36.16	110.23	13.49
3	0	70	17.9		-.34	-.31	-.05	-.50	-.45	-.07		27.59	93.14	12.86
4	0	32	8.2		-.20	-.14	-.09	-.37	-.25	-.16		28.03	96.53	12.22
TOTAL		390												

TEST NO 1 TOEFL SUBTEST 1 SECTION 1: LISTENING

ITEM NUMBER 36

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	101	25.9		-.13	-.11	-.07	-.18	-.15	-.10		31.67	101.60	12.78
2	0	108	27.7		-.06	-.01	.01	-.08	-.01	.01		32.74	104.72	13.23
C 3	1	108	27.7	C	.18	.14	.06	.24	.19	.08	C	35.84	109.12	13.49
4	0	73	18.7		.01	-.02	.00	.01	-.03	.00		33.56	104.22	13.19
TOTAL		390												

ITEM NUMBER 37

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	75	19.2		-.19	-.19	-.16	-.28	-.27	-.24		30.32	98.07	12.08
2	0	49	12.6		-.20	-.15	.01	-.32	-.24	.01		29.18	97.80	13.24
3	0	70	17.9		-.02	-.06	-.04	-.03	-.09	-.06		33.11	102.54	12.91
C 4	1	196	50.3	C	.30	.30	.15	.38	.17	.19	C	35.89	110.41	13.68
TOTAL		390												

ITEM NUMBER 38

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	12	3.1		-.11	-.13	-.04	-.28	-.32	-.10		28.33	91.83	12.42
2	0	8	2.1		-.09	-.08	.03	-.27	-.23	.07		28.25	95.00	13.75
C 3	1	364	93.3	C	.17	.18	.00	.30	.32	.00	C	33.85	105.91	13.18
4	0	6	1.5		-.09	-.09	.03	-.27	-.29	.10		28.00	91.83	14.00
TOTAL		390												

ITEM NUMBER 39

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	53	13.6		-.31	-.25	-.08	-.48	-.39	-.13		27.25	93.74	12.49
C 2	1	212	54.4	C	.52	.46	.13	.65	.57	.17	C	37.32	112.58	13.58
3	0	71	18.2		-.22	-.19	-.02	-.32	-.28	-.02		29.69	97.62	13.07
4	0	54	13.8		-.20	-.19	-.09	-.31	-.30	-.14		29.48	96.26	12.43
TOTAL		390												

ITEM NUMBER 40 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
C 1	1	277	71.0	C .42	.38	.03	.55	.50	.04 C	35.62	109.42	13.25
2	0	27	6.9	-.16	-.12	-.02	-.31	-.24	-.04	28.70	96.78	12.93
3	0	28	7.2	-.27	-.23	.03	-.51	-.43	.06	25.71	90.07	13.54
4	0	58	14.9	-.22	-.23	-.05	-.34	-.35	-.07	29.21	95.21	12.81
TOTAL		390										

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 18

TEST NO 1 TOEFL SUBTEST 1 SECTION 1: LISTENING

ITEM NUMBER 41 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	17	4.4	-.17	-.11	.05	-.37	-.24	.11	27.12	95.71	13.94
C 2	1	314	80.5	C .37	.31	.07	.53	.44	.11 C	34.94	107.78	13.30
3	0	46	11.8	-.26	-.23	-.07	-.43	-.37	-.12	27.72	93.80	12.52
4	0	13	3.3	-.15	-.15	-.09	-.37	-.35	-.22	26.77	90.85	11.62
TOTAL		390										

ITEM NUMBER 42 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	36	9.2	-.29	-.30	-.12	-.51	-.53	-.21	26.08	88.00	11.94
2	0	33	8.5	-.16	-.12	-.06	-.28	-.21	-.10	29.33	98.18	12.58
3	0	37	9.5	-.17	-.17	-.07	-.29	-.30	-.12	29.27	95.46	12.49
C 4	1	282	72.3	C .39	.38	.17	.53	.51	.23 C	35.44	109.33	13.52
OTHER	0	2	.5	.02	-.03	-.08	.09	-.15	-.39	35.50	97.00	9.50
TOTAL		390										

ITEM NUMBER 43 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
C 1	1	226	57.9	C .43	.44	.15	.54	.55	.19 C	36.41	111.81	13.60
2	0	65	16.7	-.18	-.18	-.06	-.27	-.27	-.09	30.26	97.78	12.72
3	0	55	14.1	-.19	-.19	-.08	-.29	-.30	-.12	29.73	96.45	12.56
4	0	42	10.8	-.25	-.25	-.07	-.42	-.42	-.12	27.67	91.88	12.50
OTHER	0	2	.5	-.02	-.07	-.04	-.09	-.33	-.18	31.50	88.00	11.50
TOTAL		390										

ITEM NUMBER 44 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	22	5.6	-.14	-.11	-.01	-.28	-.23	-.02	28.86	96.77	13.05
C 2	1	189	48.5	C .38	.33	.14	.48	.41	.17 C	36.66	111.14	13.65
3	0	102	26.2	-.25	-.19	-.09	-.34	-.26	-.12	30.12	99.14	12.71
4	0	76	19.5	-.12	-.13	-.07	-.17	-.19	-.10	31.51	100.21	12.70
OTHER	0	1	.3	-.05	-.01	.00	-.34	-.05	-.02	25.00	102.00	13.00
TOTAL		390										

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 19

TEST NO 1 TOEFL SUBTEST 1 SECTION 1: LISTENING

ITEM NUMBER 45 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	24	6.2	-.20	-.16	-.10	-.39	-.32	-.20	27.21	93.67	11.92
2	0	15	3.8	-.16	-.12	-.06	-.37	-.28	-.13	26.93	94.20	12.27
C 3	1	345	88.5	C .31	.25	.12	.48	.39	.19 C	34.37	106.66	13.32
4	0	5	1.3	-.12	-.11	-.01	-.42	-.36	-.05	24.80	88.20	12.80
OTHER	0	1	.3	-.10	-.11	-.05	-.62	-.70	-.32	18.00	66.00	10.00
TOTAL		390										

ITEM NUMBER 46 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
C 1	1	268	68.7	C .39	.38	.12	.52	.49	.15 C	35.62	109.63	13.43
2	0	25	6.4	-.24	-.24	-.09	-.47	-.47	-.17	26.08	88.52	12.12
3	0	86	22.1	-.22	-.19	-.05	-.31	-.27	-.08	30.09	98.57	12.85
4	0	10	2.6	-.17	-.20	-.05	-.44	-.53	-.13	25.20	82.90	12.20
OTHER	0	1	.3	-.10	-.11	-.05	-.62	-.70	-.32	18.00	66.00	10.00
TOTAL		390										

ITEM NUMBER 47

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		COEFFICIENTS OF CORRELATION ST	MEANS TT	MEANS EC
1	0	37	9.5		-.26	-.23	.00	-.46	-.41	.01		26.92	92.00	13.22
2	0	21	5.4		-.27	-.27	-.15	-.55	-.55	-.31		24.48	84.67	11.14
3	0	7	1.8		-.18	-.17	.02	-.54	-.52	.07		22.71	82.00	13.71
C 4	1	325	83.3	C	.43	.41	.08	.63	.60	.12	C	35.04	108.34	13.30
TOTAL		390												

ITEM NUMBER 48

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		COEFFICIENTS OF CORRELATION ST	MEANS TT	MEANS EC
C 1	1	95	24.4	C	.46	.44	.16	.63	.60	.22	C	40.05	119.09	14.09
2	0	31	7.9		-.19	-.15	-.03	-.34	-.28	-.06		28.32	95.61	12.84
3	0	143	36.7		-.30	-.32	-.13	-.38	-.41	-.17		30.32	97.44	12.62
4	0	121	31.0		-.01	.01	.01	-.01	.02	.01		33.36	105.40	13.21
TOTAL		390												

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 20

TEST NO 1 TOEFL SUBTEST 1 SECTION 1: LISTENING

ITEM NUMBER 49

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		COEFFICIENTS OF CORRELATION ST	MEANS TT	MEANS EC
1	0	51	13.1		-.21	-.17	-.08	-.33	-.27	-.13		29.18	97.04	12.47
2	0	23	5.9		-.16	-.19	-.02	-.33	-.39	-.05		28.17	90.96	12.87
C 3	1	267	68.5	C	.30	.28	.09	.40	.37	.12	C	35.13	108.52	13.38
4	0	49	12.6		-.10	-.09	-.02	-.16	-.14	-.04		31.41	100.98	12.98
TOTAL		390												

ITEM NUMBER 50

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		COEFFICIENTS OF CORRELATION ST	MEANS TT	MEANS EC
1	0	32	8.2		-.09	-.06	-.01	-.17	-.11	-.01		31.00	101.28	13.09
C 2	1	301	77.2	C	.45	.45	.16	.62	.61	.22	C	35.46	109.42	13.47
3	0	18	4.6		-.23	-.24	-.09	-.49	-.51	-.20		25.22	85.61	11.83
4	0	37	9.5		-.36	-.38	-.15	-.63	-.66	-.26		24.51	83.73	11.68
OTHER	0	2	.5		-.16	-.12	-.05	-.78	-.58	-.23		15.50	75.00	11.00
TOTAL		390												

1 LERTAP 2.0 SUBTEST STATISTICS
PAGE 21

TEST NO 1 TOEFL SUBTEST 1 SECTION 1: LISTENING

NUMBER OF INDIVIDUALS = 390.00 NUMBER OF ITEMS = 50.00

MEAN = 33.48 HIGHEST SCORE = 49.00

STANDARD DEVIATION = 8.06 LOWEST SCORE = 11.00

SOURCE OF VARIANCE	D.F.	S.S.	M.S.
INDIVIDUALS	389.00	505.63	1.30
ITEMS	49.00	579.32	11.82
RESIDUAL	19061.00	3229.56	.17
TOTAL	19499.00	4314.51	.22

HOYT ESTIMATE OF RELIABILITY = .87

STANDARD ERROR OF MEASUREMENT = 2.88

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 22

TEST NO 1 TOEFL SUBTEST 2 SECTION 2: STRUCTURE

ITEM NUMBER 1

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		COEFFICIENTS OF CORRELATION ST	MEANS TT	MEANS EC
1	0	7	1.8		-.01	.00	-.02	-.04	.00	-.06		31.00	104.86	12.71
2	0	9	2.3		-.29	-.19	-.08	-.78	-.52	-.22		22.44	82.78	11.44
3	0	3	.8		.06	.09	.00	.24	.37	.02		34.67	123.33	13.33
C 4	1	371	95.1	C	.18	.10	.07	.34	.18	.12	C	31.64	105.43	13.23
TOTAL		390												

ITEM NUMBER 2

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	15	3.8		-.29	-.25	-.15	-.66	-.59	-.35		24.47	82.00	10.73
C 2	1	325	83.3	C	.44	.43	.12	.64	.63	.17	C	32.19	108.50	13.35
3	0	20	5.1		-.14	-.13	.06	-.30	-.26	.12		28.45	95.30	14.00
4	0	30	7.7		-.29	-.31	-.10	-.53	-.57	-.19		26.63	85.53	12.03
TOTAL		390												

ITEM NUMBER 3

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	15	3.8		-.31	-.27	-.21	-.71	-.63	-.48		24.00	80.20	9.80
2	0	5	1.3		-.26	-.14	-.10	-.87	-.46	-.35		20.60	83.40	10.20
C 3	1	368	94.4	C	.39	.30	.21	.70	.54	.37	C	31.90	106.38	13.35
4	0	2	.5		-.03	-.02	.05	-.14	-.11	.25		29.50	99.50	15.50
TOTAL		390												

ITEM NUMBER 4

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	0	.0		.00	.00	.00	.00	.00	.00		.00	.00	.00
2	0	1	.3		-.09	-.10	-.03	-.56	-.66	-.22		23.00	68.00	11.00
3	0	9	2.3		-.34	-.19	-.11	-.93	-.52	-.30		20.78	82.78	10.89
C 4	1	380	97.4	C	.35	.21	.11	.69	.42	.22	C	31.71	105.66	13.24
TOTAL		390												

ITEM NUMBER 5

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	13	3.3		-.20	-.19	-.02	-.48	-.45	-.06		26.31	86.92	12.77
2	0	7	1.8		-.09	-.07	.03	-.26	-.22	.08		28.29	95.29	13.86
C 3	1	351	90.0	C	.31	.27	.07	.50	.44	.12	C	31.93	106.67	13.26
4	0	19	4.9		-.21	-.18	-.10	-.45	-.38	-.21		26.95	90.84	11.74
TOTAL		390												

TEST NO 1 TOEFL SUBTEST 2 SECTION 2: STRUCTURE

ITEM NUMBER 6

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	2	.5		-.06	-.08	-.03	-.28	-.37	-.13		27.50	85.50	12.00
2	0	69	17.7		-.26	-.31	-.15	-.39	-.45	-.22		28.70	93.10	12.13
C 3	1	319	81.8	C	.27	.32	.15	.39	.46	.22	C	32.06	107.74	13.41
4	0	0	.0		.00	.00	.00	.00	.00	.00		.00	.00	.00
TOTAL		390												

ITEM NUMBER 7

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
C 1	1	366	93.8	C	.35	.30	.13	.63	.54	.22	C	31.88	106.45	13.28
2	0	2	.5		-.04	-.08	-.06	-.21	-.37	-.29		28.50	86.00	10.50
3	0	17	4.4		-.31	-.27	-.09	-.68	-.60	-.20		24.47	81.88	11.76
4	0	5	1.3		-.17	-.11	-.06	-.58	-.37	-.21		24.20	88.00	11.40
TOTAL		390												

ITEM NUMBER 8

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	43	11.0		-.20	-.21	-.08	-.34	-.35	-.14		28.63	94.30	12.40
2	0	19	4.9		-.21	-.21	-.07	-.45	-.45	-.15		26.89	88.26	12.16
3	0	27	6.9		-.35	-.25	-.14	-.66	-.47	-.26		25.26	88.74	11.56
C 4	1	301	77.2	C	.47	.41	.18	.65	.57	.25	C	32.68	109.09	13.50
TOTAL		390												

ITEM NUMBER 9

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	12	3.1		-.04	-.06	.02	-.11	-.14	.04		30.25	99.17	13.50
2	0	41	10.5		-.24	-.24	-.20	-.40	-.41	-.34		28.07	92.34	11.24
3	0	3	.8		-.23	-.19	-.08	-.93	-.80	-.32		19.00	65.33	10.33
C 4	1	334	85.6	C	.29	.29	.19	.43	.43	.28	C	32.01	107.16	13.43
TOTAL		390												

ITEM NUMBER 10 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
C 1	1	312	80.0	C	.33	.32	.11	.46	.46	.16	C	32.23	107.97	13.37
2	0	20	5.1		-.23	-.20	-.09	-.47	-.41	-.19		26.75	89.85	11.90
3	0	43	11.0		-.20	-.24	-.07	-.33	-.40	-.11		28.72	92.81	12.56
4	0	13	3.3		-.11	-.05	-.01	-.26	-.12	-.02		28.62	100.31	13.00
OTHER	0	2	.5		.00	-.05	-.04	.00	-.22	-.18		31.50	93.50	11.50
TOTAL		390												

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 24

TEST NO 1 TOEFL SUBTEST 2 SECTION 2: STRUCTURE

ITEM NUMBER 11 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	2	.5		-.17	-.04	-.07	-.82	-.21	-.34		20.00	94.00	10.00
C 2	1	282	72.3	C	.38	.31	.14	.51	.42	.19	C	32.59	108.54	13.46
3	0	0	.0		.00	.00	.00	.00	.00	.00		.00	.00	.00
4	0	106	27.2		-.36	-.31	-.13	-.48	-.41	-.18		28.59	95.92	12.48
TOTAL		390												

ITEM NUMBER 12 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	57	14.6		-.38	-.32	-.25	-.59	-.50	-.38		26.96	90.95	11.25
2	0	5	1.3		-.17	-.05	-.01	-.58	-.16	-.05		24.20	97.80	12.80
C 3	1	321	82.3	C	.46	.36	.25	.67	.52	.37	C	32.47	108.05	13.56
4	0	6	1.5		-.16	-.12	-.05	-.52	-.39	-.14		25.17	87.17	12.00
OTHER	0	1	.3		-.04	-.05	-.05	-.23	-.34	-.32		28.00	86.00	10.00
TOTAL		390												

ITEM NUMBER 13 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
C 1	1	273	70.0	C	.46	.38	.12	.60	.50	.15	C	32.89	109.51	13.43
2	0	6	1.5		-.16	-.13	-.08	-.52	-.41	-.25		25.17	86.33	11.17
3	0	44	11.3		-.32	-.32	-.12	-.53	-.53	-.20		27.07	88.77	12.09
4	0	67	17.2		-.23	-.15	-.02	-.35	-.22	-.02		28.96	99.16	13.06
TOTAL		390												

ITEM NUMBER 14 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	6	1.5		-.22	-.17	-.10	-.71	-.54	-.31		22.83	80.50	10.67
C 2	1	293	75.1	C	.25	.20	.03	.34	.27	.04	C	32.14	107.11	13.24
3	0	83	21.3		-.14	-.11	-.02	-.20	-.15	-.02		30.11	101.33	13.08
4	0	8	2.1		-.16	-.15	.03	-.46	-.43	.09		26.13	86.13	13.88
TOTAL		390												

ITEM NUMBER 15 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
C 1	1	96	24.6	C	.35	.30	.18	.47	.41	.25	C	34.38	114.60	14.21
2	0	16	4.1		-.09	-.10	-.10	-.21	-.22	-.23		29.25	96.50	11.56
3	0	151	38.7		-.16	-.16	-.10	-.20	-.21	-.13		30.46	101.36	12.75
4	0	127	32.6		-.11	-.07	-.01	-.15	-.09	-.02		30.65	103.25	13.11
TOTAL		390												

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 25

TEST NO 1 TOEFL SUBTEST 2 SECTION 2: STRUCTURE

ITEM NUMBER 16 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	12	3.1		-.11	-.05	-.11	-.28	-.12	-.26		28.33	100.08	11.25
C 2	1	376	96.4	C	.19	.09	.12	.36	.17	.24	C	31.61	105.34	13.26
3	0	2	.5		-.21	-.11	-.07	-1.03	-.52	-.34		17.00	78.00	10.00
4	0	0	.0		.00	.00	.00	.00	.00	.00		.00	.00	.00
TOTAL		390												

ITEM NUMBER 17 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	2	.5		-.01	-.01	-.03	-.03	-.07	-.13		31.00	101.50	12.00
C 2	1	357	91.5	C	.20	.18	.07	.33	.30	.11	C	31.73	106.01	13.25
3	0	22	5.6		-.09	-.07	-.04	-.19	-.14	-.08		29.64	100.09	12.68
4	0	9	2.3		-.23	-.22	-.06	-.62	-.60	-.15		24.33	79.33	12.00
TOTAL		390												

ITEM NUMBER 18 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
C 1	1	361	92.6	C	.38	.31	.09	.65	.53	.16	C	31.96	106.63	13.27
2	0	9	2.3		-.20	-.16	.00	-.55	-.45	-.01		25.11	85.67	13.11
3	0	11	2.8		-.21	-.18	-.04	-.55	-.45	-.11		25.36	86.36	12.36
4	0	9	2.3		-.23	-.19	-.11	-.62	-.51	-.31		24.33	83.22	10.78
TOTAL		390												

ITEM NUMBER 19 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	6	1.5		-.23	-.09	-.05	-.73	-.29	-.16		22.50	91.83	11.83
2	0	8	2.1		-.08	-.04	-.05	-.22	-.13	-.13		28.88	99.50	12.13
3	0	10	2.6		-.20	-.19	.07	-.52	-.51	.17		25.60	83.60	14.50
C 4	1	366	93.8	C	.29	.20	.01	.52	.35	.02	C	31.80	105.96	13.19
TOTAL		390												

ITEM NUMBER 20 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
C 1	1	356	91.3	C	.26	.17	.12	.44	.28	.20	C	31.83	105.97	13.30
2	0	13	3.3		-.07	-.07	-.06	-.18	-.18	-.14		29.54	97.92	12.15
3	0	9	2.3		-.26	-.12	-.10	-.71	-.33	-.28		23.33	90.89	11.00
4	0	12	3.1		-.13	-.09	-.04	-.33	-.23	-.10		27.83	95.67	12.42
TOTAL		390												

ITEM NUMBER 21 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	7	1.8		-.18	-.14	-.04	-.54	-.42	-.13		25.00	86.43	12.14
2	0	57	14.6		-.39	-.38	-.15	-.60	-.59	-.22		26.91	88.33	12.04
3	0	4	1.0		-.09	-.07	-.01	-.33	-.27	-.05		27.25	92.00	12.75
C 4	1	322	82.6	C	.45	.42	.15	.65	.62	.22	C	32.43	108.56	13.41
TOTAL		390												

ITEM NUMBER 22 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	23	5.9		-.22	-.23	-.09	-.45	-.46	-.19		27.09	88.52	11.96
2	0	23	5.9		-.26	-.23	-.10	-.52	-.46	-.21		26.39	88.61	11.83
3	0	14	3.6		-.21	-.15	-.11	-.50	-.35	-.26		26.07	91.29	11.29
C 4	1	330	84.6	C	.43	.37	.19	.64	.56	.28	C	32.32	107.92	13.44
TOTAL		390												

ITEM NUMBER 23 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	16	4.1		-.29	-.26	-.12	-.64	-.59	-.27		24.75	82.00	11.31
2	0	5	1.3		-.09	-.10	-.02	-.31	-.33	-.07		27.60	89.60	12.60
C 3	1	366	93.8	C	.32	.26	.11	.57	.47	.19	C	31.84	106.27	13.27
4	0	3	.8		-.12	-.01	.00	-.48	-.02	.02		25.00	104.00	13.33
TOTAL		390												

ITEM NUMBER 24 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	4	1.0		-.04	-.02	.02	-.13	-.06	.07		29.75	102.00	13.75
2	0	11	2.8		-.27	-.17	-.09	-.69	-.43	-.23		23.82	87.09	11.45
C 3	1	352	90.3	C	.39	.29	.09	.63	.47	.15	C	32.06	106.76	13.28
4	0	23	5.9		-.29	-.24	-.06	-.57	-.48	-.11		25.91	87.87	12.43
TOTAL		390												

ITEM NUMBER 25 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	15	3.8		-.19	-.18	-.06	-.45	-.41	-.13		26.73	88.87	12.27
C 2	1	361	92.6	C	.32	.27	.06	.55	.47	.10	C	31.88	106.45	13.24
3	0	2	.5		-.15	-.03	.05	-.71	-.13	.25		21.50	98.50	15.50
4	0	12	3.1		-.21	-.21	-.05	-.52	-.52	-.13		25.75	84.00	12.25
TOTAL		390												

TEST NO 1 TOEFL SUBTEST 2 SECTION 2: STRUCTURE

ITEM NUMBER 26 COEFFICIENTS OF CORRELATION MEANS

OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	47	12.1	-.25	-.26	-.17	-.41	-.42	-.28	28.17	92.53	11.64
C 2	1	315	80.8	C .46	.43	.23	.65	.62	.33 C	32.52	108.86	13.55
3	0	10	2.6	-.16	-.16	-.06	-.41	-.42	-.17	26.80	87.50	11.90
4	0	18	4.6	-.35	-.30	-.11	-.77	-.64	-.25	23.67	80.61	11.50
TOTAL		390										

ITEM NUMBER 27 COEFFICIENTS OF CORRELATION MEANS

OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
C 1	1	330	84.6	C .29	.24	.12	.43	.35	.18 C	32.04	106.87	13.35
2	0	14	3.6	-.21	-.16	-.09	-.50	-.39	-.21	26.14	89.71	11.64
3	0	20	5.1	-.16	-.15	-.06	-.33	-.32	-.12	28.15	93.30	12.40
4	0	26	6.7	-.12	-.09	-.06	-.24	-.17	-.12	29.23	99.04	12.42
TOTAL		390										

ITEM NUMBER 28 COEFFICIENTS OF CORRELATION MEANS

OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	14	3.6	-.38	-.30	-.18	-.89	-.70	-.41	22.00	77.29	10.21
C 2	1	362	92.8	C .43	.38	.20	.74	.65	.34 C	32.02	106.94	13.36
3	0	6	1.5	-.21	-.24	-.12	-.67	-.78	-.37	23.33	69.67	10.17
4	0	6	1.5	-.06	-.07	.01	-.20	-.21	.04	29.00	95.50	13.50
OTHER	0	2	.5	-.10	-.06	-.08	-.50	-.28	-.39	24.50	90.50	9.50
TOTAL		390										

ITEM NUMBER 29 COEFFICIENTS OF CORRELATION MEANS

OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	6	1.5	-.14	-.11	-.11	-.46	-.33	-.35	25.83	89.83	10.33
2	0	100	25.6	-.37	-.26	-.13	-.51	-.35	-.18	28.36	97.05	12.45
C 3	1	281	72.1	C .42	.29	.15	.56	.39	.20 C	32.71	108.33	13.49
4	0	2	.5	-.04	-.05	.08	-.21	-.25	.41	28.50	92.00	17.00
OTHER	0	1	.3	-.11	-.03	-.06	-.70	-.18	-.41	21.00	95.00	9.00
TOTAL		390										

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 28

TEST NO 1 TOEFL SUBTEST 2 SECTION 2: STRUCTURE

ITEM NUMBER 30 COEFFICIENTS OF CORRELATION MEANS

OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	7	1.8	-.15	-.13	-.04	-.44	-.38	-.13	26.14	88.00	12.14
C 2	1	319	81.8	C .43	.40	.08	.62	.58	.12 C	32.43	108.49	13.30
3	0	32	8.2	-.28	-.28	-.02	-.50	-.50	-.04	26.97	88.28	12.94
4	0	30	7.7	-.24	-.22	-.05	-.45	-.40	-.09	27.37	91.43	12.60
OTHER	0	2	.5	-.10	-.07	-.08	-.50	-.36	-.39	24.50	86.50	9.50
TOTAL		390										

ITEM NUMBER 31 COEFFICIENTS OF CORRELATION MEANS

OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	22	5.6	-.16	-.15	-.10	-.33	-.30	-.21	28.27	94.18	11.82
2	0	41	10.5	-.43	-.36	-.11	-.72	-.62	-.19	25.37	85.78	12.10
C 3	1	320	82.1	C .50	.44	.19	.72	.63	.27 C	32.57	108.74	13.46
4	0	7	1.8	-.18	-.17	-.10	-.53	-.51	-.29	25.14	82.57	10.86
TOTAL		390										

ITEM NUMBER 32 COEFFICIENTS OF CORRELATION MEANS

OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
C 1	1	267	68.5	C .28	.26	.11	.37	.33	.15 C	32.17	108.18	13.43
2	0	8	2.1	-.10	-.07	-.04	-.27	-.21	-.10	28.25	96.00	12.38
3	0	36	9.2	-.24	-.19	-.15	-.42	-.33	-.26	27.81	94.36	11.67
4	0	79	20.3	-.12	-.13	-.01	-.17	-.19	-.02	30.27	100.20	13.10
TOTAL		390										

ITEM NUMBER 33 COEFFICIENTS OF CORRELATION MEANS

OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	37	9.5	-.07	-.04	.01	-.12	-.07	.01	30.38	102.92	13.24
C 2	1	281	72.1	C .18	.14	.02	.24	.19	.03 C	31.99	106.63	13.22
3	0	60	15.4	-.09	-.09	-.01	-.13	-.14	-.01	30.47	101.23	13.13
4	0	12	3.1	-.17	-.11	-.06	-.43	-.28	-.14	26.75	93.42	12.17
TOTAL		390										

ITEM NUMBER 34 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	10	2.6		-.28	-.22	-.05	-.74	-.59	-.13		23.10	80.30	12.20
C 2	1	270	69.2	C	.24	.19	.03	.31	.24	.04	C	32.21	107.28	13.24
3	0	48	12.3		-.06	-.06	.05	-.09	-.10	.09		30.71	102.19	13.65
4	0	61	15.6		-.12	-.08	-.06	-.18	-.12	-.09		30.07	101.77	12.75
OTHER	0	1	.3		-.07	-.06	-.05	-.43	-.38	-.32		25.00	84.00	10.00
TOTAL		390												

TEST NO 1 TOEFL — SUBTEST 2 SECTION 2: STRUCTURE

ITEM NUMBER 35 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	12	3.1		-.01	-.02	.00	-.03	-.06	.01		31.08	102.75	13.25
2	0	12	3.1		-.27	-.24	-.08	-.67	-.59	-.21		24.17	81.08	11.67
3	0	110	28.2		-.14	-.09	-.06	-.19	-.12	-.08		30.33	102.50	12.85
C 4	1	256	65.6	C	.24	.18	.09	.31	.23	.11	C	32.27	107.36	13.39
TOTAL		390												

ITEM NUMBER 36 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
C 1	1	231	59.2	C	.29	.27	.11	.37	.34	.14	C	32.61	109.06	13.48
2	0	101	25.9		-.03	-.03	-.01	-.04	-.04	-.02		31.21	104.23	13.10
3	0	26	6.7		-.02	-.05	.01	-.04	-.11	.02		31.04	101.35	13.31
4	0	32	8.2		-.46	-.39	-.18	-.82	-.70	-.33		24.06	81.56	11.19
TOTAL		390												

ITEM NUMBER 37 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
C 1	1	240	61.5	C	.35	.30	.13	.44	.39	.16	C	32.78	109.39	13.51
2	0	53	13.6		-.19	-.15	-.02	-.30	-.24	-.03		29.09	98.11	13.02
3	0	19	4.9		-.25	-.23	-.12	-.54	-.49	-.26		26.05	86.63	11.42
4	0	78	20.0		-.12	-.12	-.07	-.18	-.17	-.10		30.23	100.85	12.71
TOTAL		390												

ITEM NUMBER 38 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	25	6.4		-.26	-.19	-.09	-.50	-.38	-.18		26.68	91.60	12.00
2	0	17	4.4		-.12	-.10	-.05	-.27	-.23	-.11		28.65	96.24	12.41
3	0	76	19.5		-.26	-.23	-.10	-.38	-.33	-.15		28.86	96.54	12.50
C 4	1	272	69.7	C	.42	.35	.16	.55	.46	.21	C	32.77	109.20	13.53
TOTAL		390												

ITEM NUMBER 39 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	26	6.7		.00	.03	-.03	.01	.07	-.05		31.50	107.35	12.85
2	0	70	17.9		-.10	-.07	-.07	-.14	-.10	-.10		30.44	102.43	12.70
3	0	91	23.3		-.24	-.19	-.15	-.32	-.27	-.21		29.37	98.69	12.26
C 4	1	203	52.1	C	.27	.20	.20	.34	.25	.25	C	32.70	108.49	13.80
TOTAL		390												

TEST NO 1 TOEFL — SUBTEST 2 SECTION 2: STRUCTURE

ITEM NUMBER 40 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	81	20.8		-.02	.02	-.04	-.03	.03	-.06		31.26	105.80	12.91
C 2	1	75	19.2	C	.04	-.03	.07	.05	-.05	.10	C	31.80	103.85	13.63
3	0	179	45.9		.02	.04	.00	.02	.05	.00		31.54	105.79	13.17
4	0	55	14.1		-.05	-.04	-.02	-.07	-.07	-.03		30.89	103.07	13.00
TOTAL		390												

TEST NO 1 TOEFL SUBTEST 2 SECTION 2: STRUCTURE

NUMBER OF INDIVIDUALS = 390.00 NUMBER OF ITEMS = 40.00

MEAN = 31.44 HIGHEST SCORE = 40.00

STANDARD DEVIATION = 4.85 LOWEST SCORE = 9.00

SOURCE OF VARIANCE	D.F.	S.S.	M.S.
INDIVIDUALS	389.00	228.70	.59
ITEMS	39.00	458.27	11.75
RESIDUAL	15171.00	1937.36	.13
TOTAL	15599.00	2624.33	.17

HOYT ESTIMATE OF RELIABILITY = .78

STANDARD ERROR OF MEASUREMENT = 2.23

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 32

TEST NO 1 TOEFL SUBTEST 3 SECTION 3: READING

ITEM NUMBER 1 COEFFICIENTS OF CORRELATION MEANS

	OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
	1	0	50	12.8	-.17	-.16	-.04	-.27	-.26	-.07	36.78	97.32	12.80
	2	0	9	2.3	-.33	-.25	-.16	-.92	-.70	-.44	23.89	75.11	9.78
	3	0	20	5.1	-.16	-.18	-.02	-.34	-.37	-.03	34.85	91.35	12.95
C	4	1	311	79.7	.36	.33	.11	.50	.46	.15	41.47	108.03	13.35
	TOTAL		390										

ITEM NUMBER 2 COEFFICIENTS OF CORRELATION MEANS

	OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
	1	0	53	13.6	-.27	-.25	-.13	-.43	-.40	-.20	35.00	93.47	12.15
C	2	1	333	85.4	.32	.29	.11	.48	.44	.17	41.11	107.21	13.33
	3	0	4	1.0	-.20	-.15	.03	-.72	-.57	.09	25.75	77.75	14.00
	4	0	0	.0	.00	.00	.00	.00	.00	.00	.00	.00	.00
	TOTAL		390										

ITEM NUMBER 3 COEFFICIENTS OF CORRELATION MEANS

	OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
	1	0	7	1.8	-.26	-.24	-.11	-.79	-.72	-.34	25.57	73.14	10.43
C	2	1	321	82.3	.46	.44	.16	.67	.64	.23	41.73	108.74	13.42
	3	0	15	3.8	-.21	-.22	-.15	-.49	-.51	-.36	32.13	84.93	10.67
	4	0	47	12.1	-.31	-.29	-.05	-.50	-.47	-.08	33.89	90.89	12.77
	TOTAL		390										

ITEM NUMBER 4 COEFFICIENTS OF CORRELATION MEANS

	OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
	1	0	69	17.7	-.22	-.17	-.06	-.32	-.26	-.09	36.58	98.23	12.77
	2	0	14	3.6	-.10	-.15	.00	-.23	-.35	.00	36.36	91.00	13.21
	3	0	12	3.1	-.20	-.16	-.05	-.51	-.41	-.11	31.58	88.50	12.33
C	4	1	295	75.6	.32	.29	.07	.44	.39	.09	41.48	107.97	13.31
	TOTAL		390										

ITEM NUMBER 5 COEFFICIENTS OF CORRELATION MEANS

	OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
	1	0	18	4.6	-.13	-.05	-.05	-.29	-.11	-.10	35.61	100.94	12.50
	2	0	36	9.2	-.31	-.29	-.10	-.53	-.51	-.18	32.97	88.61	12.14
C	3	1	266	68.2	.46	.41	.20	.60	.53	.26	42.48	110.07	13.62
	4	0	70	17.9	-.26	-.25	-.14	-.38	-.36	-.20	36.00	95.43	12.21
	TOTAL		390										

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 33

TEST NO 1 TOEFL SUBTEST 3 SECTION 3: READING

ITEM NUMBER 6

				COEFFICIENTS OF CORRELATION						MEANS		
OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	36	9.2	-.31	-.30	-.08	-.55	-.52	-.14	32.81	88.28	12.39
C 2	1	330	84.6 C	.47	.46	.20	.70	.68	.29 C	41.61	108.55	13.45
3	0	16	4.1	-.32	-.30	-.21	-.72	-.69	-.46	28.56	78.38	9.94
4	0	8	2.1	-.10	-.13	-.06	-.30	-.37	-.17	34.75	89.00	11.88
TOTAL		390										

ITEM NUMBER 7

				COEFFICIENTS OF CORRELATION						MEANS		
OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
C 1	1	233	59.7 C	.35	.37	.21	.44	.47	.27 C	42.25	110.54	13.75
2	0	8	2.1	-.14	-.13	-.02	-.39	-.37	-.07	33.13	88.75	12.63
3	0	75	19.2	-.25	-.27	-.17	-.37	-.39	-.25	36.24	95.13	12.03
4	0	74	19.0	-.13	-.15	-.08	-.19	-.21	-.12	38.11	99.51	12.61
TOTAL		390										

ITEM NUMBER 8

				COEFFICIENTS OF CORRELATION						MEANS		
OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	1	.3	-.05	-.04	-.03	-.31	-.27	-.22	33.00	90.00	11.00
2	0	23	5.9	-.29	-.22	-.10	-.59	-.43	-.20	31.35	89.39	11.87
C 3	1	350	89.7 C	.27	.21	.10	.44	.33	.17 C	40.81	106.29	13.29
4	0	16	4.1	-.05	-.05	-.03	-.12	-.10	-.07	38.25	101.00	12.69
TOTAL		390										

ITEM NUMBER 9

				COEFFICIENTS OF CORRELATION						MEANS		
OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	24	6.2	-.21	-.19	-.12	-.42	-.37	-.23	33.96	91.96	11.71
2	0	77	19.7	-.24	-.24	-.16	-.35	-.34	-.23	36.49	96.26	12.12
3	0	18	4.6	-.26	-.28	-.06	-.57	-.60	-.12	31.17	82.22	12.33
C 4	1	270	69.2 C	.45	.44	.24	.59	.58	.31 C	42.37	110.37	13.70
OTHER	0	1	.3	-.12	-.11	-.11	-.78	-.72	-.71	22.00	65.00	6.00
TOTAL		390										

ITEM NUMBER 10

				COEFFICIENTS OF CORRELATION						MEANS		
OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	2	.5	-.12	-.12	.01	-.56	-.60	.03	28.00	74.00	13.50
2	0	2	.5	-.14	-.09	-.07	-.68	-.44	-.34	25.50	82.00	10.00
C 3	1	384	98.5 C	.22	.20	.07	.45	.40	.14 C	40.33	105.49	13.21
4	0	2	.5	-.13	-.13	-.06	-.61	-.61	-.29	27.00	73.00	10.50
TOTAL		390										

TEST NO 1 TOEFL SUBTEST 3 SECTION 3: READING

ITEM NUMBER 11

				COEFFICIENTS OF CORRELATION						MEANS		
OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	3	.8	.01	-.02	-.02	.03	-.09	-.09	40.67	100.33	12.33
C 2	1	354	90.8 C	.27	.24	.18	.44	.19	.30 C	40.76	106.41	13.37
3	0	16	4.1	-.24	-.23	-.12	-.55	-.53	-.27	31.31	84.50	11.31
4	0	17	4.4	-.15	-.10	-.13	-.32	-.22	-.29	35.00	96.59	11.18
TOTAL		390										

ITEM NUMBER 12

				COEFFICIENTS OF CORRELATION						MEANS		
OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
C 1	1	375	96.2 C	.31	.29	-.01	.60	.56	-.03 C	40.59	106.10	13.17
2	0	4	1.0	-.12	-.11	-.11	-.43	-.39	-.40	31.50	86.25	9.75
3	0	3	.8	-.11	-.11	.15	-.44	-.44	.61	31.00	83.00	18.67
4	0	8	2.1	-.28	-.26	.00	-.79	-.74	.01	25.88	73.00	13.25
TOTAL		390										

ITEM NUMBER 13

				COEFFICIENTS OF CORRELATION						MEANS		
OPTION	WT	N	P	PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC	ST	TT	EC
1	0	12	3.1	-.21	-.21	-.01	-.51	-.51	-.01	31.50	84.08	13.08
2	0	35	9.0	-.44	-.41	-.26	-.77	-.72	-.46	29.71	81.57	10.49
C 3	1	277	71.0 C	.47	.42	.14	.62	.56	.19 C	42.37	109.90	13.47
4	0	66	16.9	-.14	-.10	.03	-.21	-.15	.04	37.79	100.91	13.39
TOTAL		390										

ITEM NUMBER 14

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	10	2.6		-.28	-.25	-.16	-.74	-.67	-.42		27.30	76.90	10.00
2	0	3	.8		-.17	-.15	-.06	-.72	-.61	-.24		25.33	74.67	11.00
3	0	2	.5		-.11	-.07	-.01	-.54	-.36	-.07		28.50	86.50	12.50
C 4	1	375	96.2	C	.35	.30	.16	.66	.57	.31	C	40.65	106.13	13.29
TOTAL		390												

ITEM NUMBER 15

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
C 1	1	324	83.1	C	.51	.46	.19	.75	.68	.28	C	41.86	108.83	13.46
2	0	16	4.1		-.28	-.23	-.08	-.63	-.53	-.19		30.00	84.69	11.88
3	0	40	10.3		-.33	-.32	-.12	-.56	-.55	-.20		32.78	87.65	12.05
4	0	10	2.6		-.23	-.19	-.12	-.61	-.49	-.31		29.50	84.40	10.80
TOTAL		390												

TEST NO 1 TOEFL SUBTEST 3 SECTION 3: READING

ITEM NUMBER 16

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	18	4.6		-.35	-.32	-.14	-.76	-.69	-.30		28.22	78.89	11.11
2	0	10	2.6		-.16	-.10	-.05	-.43	-.26	-.13		32.70	94.20	12.20
3	0	9	2.3		-.18	-.13	-.06	-.48	-.36	-.17		31.56	89.44	11.89
C 4	1	353	90.5	C	.43	.35	.16	.70	.57	.26	C	41.16	107.08	13.35
TOTAL		390												

ITEM NUMBER 17

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
C 1	1	376	96.4	C	.29	.26	.11	.55	.50	.21	C	40.54	105.95	13.25
2	0	5	1.3		-.18	-.17	-.06	-.61	-.56	-.19		28.40	78.80	11.60
3	0	4	1.0		-.20	-.20	-.11	-.74	-.74	-.43		25.50	69.75	9.50
4	0	5	1.3		-.12	-.09	-.02	-.41	-.31	-.07		32.20	90.60	12.60
TOTAL		390												

ITEM NUMBER 18

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	11	2.8		-.22	-.18	-.04	-.56	-.46	-.11		30.45	86.09	12.36
C 2	1	363	93.1	C	.38	.33	.15	.67	.58	.25	C	40.91	106.68	13.31
3	0	6	1.5		-.12	-.11	-.10	-.39	-.35	-.33		32.83	89.00	10.50
4	0	10	2.6		-.29	-.26	-.11	-.77	-.69	-.29		26.70	75.90	11.00
TOTAL		390												

ITEM NUMBER 19

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
C 1	1	316	81.0	C	.46	.42	.18	.65	.61	.26	C	41.78	108.76	13.46
2	0	11	2.8		-.31	-.29	-.11	-.78	-.74	-.28		26.73	74.45	11.09
3	0	34	8.7		-.28	-.26	-.07	-.49	-.47	-.13		33.47	89.62	12.41
4	0	29	7.4		-.19	-.17	-.12	-.36	-.32	-.22		35.00	94.21	11.79
TOTAL		390												

ITEM NUMBER 20

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	18	4.6		-.30	-.24	-.13	-.66	-.52	-.28		29.78	85.22	11.28
C 2	1	362	92.8	C	.32	.27	.13	.55	.46	.23	C	40.78	106.39	13.30
3	0	9	2.3		-.08	-.09	-.03	-.23	-.23	-.08		36.11	95.00	12.56
4	0	1	.3		-.12	-.12	-.06	-.74	-.75	-.41		23.00	63.00	9.00
TOTAL		390												

TEST NO 1 TOEFL SUBTEST 3 SECTION 3: READING

ITEM NUMBER 21

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	10	2.6		-.28	-.25	-.07	-.74	-.67	-.18		27.30	76.70	11.80
2	0	13	3.3		-.24	-.20	-.19	-.58	-.49	-.45		30.54	85.23	9.92
3	0	20	5.1		-.25	-.23	-.10	-.53	-.49	-.21		32.00	86.95	11.75
C 4	1	347	89.0	C	.45	.41	.21	.72	.65	.34	C	41.32	107.64	13.42
TOTAL		390												

ITEM NUMBER 22

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEAN	TT MEAN	EC MEAN
C 1	1	239	61.3	C	.58	.51	.24	.74	.65	.31	C	43.57	112.39	13.80
2	0	42	10.8		-.17	-.11	-.06	-.29	-.19	-.11		36.38	99.17	12.57
3	0	45	11.5		-.32	-.27	-.14	-.53	-.44	-.23		33.44	91.47	11.89
4	0	64	16.4		-.34	-.34	-.14	-.51	-.52	-.21		34.41	90.98	12.16
TOTAL		390												

ITEM NUMBER 23

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEAN	TT MEAN	EC MEAN
1	0	8	2.1		-.17	-.14	-.07	-.49	-.40	-.20		31.25	87.63	11.63
2	0	9	2.3		-.26	-.23	-.08	-.70	-.62	-.22		27.67	78.44	11.44
3	0	24	6.2		-.29	-.29	-.09	-.57	-.57	-.17		31.75	84.71	12.08
C 4	1	349	89.5	C	.43	.40	.14	.69	.64	.22	C	41.22	107.52	13.34
TOTAL		390												

ITEM NUMBER 24

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEAN	TT MEAN	EC MEAN
C 1	1	290	74.4	C	.39	.37	.19	.53	.50	.26	C	41.86	108.99	13.55
2	0	63	16.2		-.25	-.22	-.11	-.38	-.33	-.16		35.86	96.03	12.38
3	0	12	3.1		-.23	-.23	-.11	-.58	-.57	-.27		30.33	81.83	11.17
4	0	25	6.4		-.16	-.17	-.10	-.32	-.34	-.20		35.48	93.04	11.88
TOTAL		390												

ITEM NUMBER 25

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEAN	TT MEAN	EC MEAN
1	0	75	19.2		-.26	-.27	-.12	-.38	-.38	-.17		36.09	95.20	12.40
2	0	86	22.1		-.12	-.09	-.03	-.16	-.12	-.04		38.48	102.00	13.00
C 3	1	208	53.3	C	.40	.39	.15	.50	.48	.19	C	42.90	111.56	13.63
4	0	21	5.4		-.20	-.22	-.07	-.42	-.46	-.14		33.76	88.05	12.24
TOTAL		390												

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 37

TEST NO 1 TOEFL SUBTEST 3 SECTION 3: READING

ITEM NUMBER 26

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEAN	TT MEAN	EC MEAN
1	0	75	19.2		-.25	-.22	-.02	-.36	-.31	-.03		36.29	96.95	13.04
2	0	58	14.9		-.20	-.17	-.08	-.31	-.27	-.13		36.48	97.52	12.53
3	0	17	4.4		-.25	-.23	-.08	-.55	-.51	-.19		31.35	85.59	11.88
C 4	1	240	61.5	C	.46	.40	.11	.58	.51	.14	C	42.82	110.76	13.47
TOTAL		390												

ITEM NUMBER 27

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEAN	TT MEAN	EC MEAN
1	0	92	23.6		-.34	-.33	-.20	-.47	-.46	-.27		35.52	94.26	12.03
2	0	2	.5		.03	.00	.00	.13	-.01	-.02		43.00	104.50	13.00
C 3	1	276	70.8	C	.42	.39	.23	.56	.51	.30	C	42.16	109.55	13.66
4	0	20	5.1		-.23	-.16	-.10	-.47	-.34	-.21		32.85	92.45	11.80
TOTAL		390												

ITEM NUMBER 28

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEAN	TT MEAN	EC MEAN
C 1	1	259	66.4	C	.38	.35	.16	.49	.46	.20	C	42.13	109.59	13.54
2	0	45	11.5		-.28	-.26	-.11	-.46	-.42	-.19		34.27	92.18	12.16
3	0	51	13.1		-.15	-.17	-.06	-.25	-.27	-.10		37.14	97.20	12.65
4	0	35	9.0		-.12	-.10	-.06	-.22	-.17	-.10		37.17	99.34	12.57
TOTAL		390												

ITEM NUMBER 29

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEAN	TT MEAN	EC MEAN
1	0	11	2.8		-.13	-.13	-.06	-.33	-.33	-.16		34.45	91.55	12.00
C 2	1	277	71.0	C	.35	.30	.08	.47	.40	.11	C	41.81	108.55	13.35
3	0	39	10.0		-.21	-.16	-.13	-.36	-.28	-.22		35.41	96.26	11.92
4	0	63	16.2		-.21	-.19	.03	-.31	-.28	.05		36.62	97.38	13.41
TOTAL		390												

ITEM NUMBER 30

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	17	4.4		-.30	-.29	-.02	-.66	-.64	-.05		29.71	80.65	12.82
2	0	35	9.0		-.38	-.35	-.16	-.68	-.62	-.28		31.00	84.86	11.54
3	0	76	19.5		-.17	-.15	-.04	-.24	-.21	-.06		37.57	99.59	12.92
C 4	1	261	66.9	C	.52	.47	.14	.67	.61	.18	C	42.83	111.00	13.51
OTHER	0	1	.3		-.11	-.06	-.03	-.70	-.38	-.22		24.00	84.00	11.00
TOTAL		390												

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 38
TEST NO 1 TOEFL SUBTEST 3 SECTION 3: READING

ITEM NUMBER 31

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
C 1	1	345	88.5	C	.48	.46	.21	.76	.72	.33	C	41.42	108.02	13.43
2	0	9	2.3		-.22	-.18	-.07	-.60	-.50	-.20		29.56	83.44	11.67
3	0	8	2.1		-.28	-.26	-.18	-.79	-.75	-.52		25.88	72.38	9.13
4	0	26	6.7		-.32	-.30	-.10	-.61	-.58	-.19		31.31	84.58	11.96
OTHER	0	2	.5		-.05	-.08	-.09	-.24	-.39	-.44		35.00	84.50	9.00
TOTAL		390												

ITEM NUMBER 32

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	4	1.0		-.14	-.13	.01	-.54	-.47	.04		29.50	82.25	13.50
2	0	13	3.3		-.20	-.21	-.11	-.49	-.50	-.26		32.00	84.85	11.31
3	0	1	.3		-.19	-.14	-.06	-1.22	-.91	-.41		12.00	54.00	9.00
C 4	1	372	95.4	C	.29	.27	.10	.53	.51	.19	C	40.60	106.13	13.25
TOTAL		390												

ITEM NUMBER 33

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	54	13.8		-.27	-.25	-.13	-.43	-.39	-.20		35.00	93.87	12.15
2	0	23	5.9		-.18	-.17	-.12	-.37	-.33	-.25		34.65	93.09	11.57
C 3	1	289	74.1	C	.52	.46	.20	.70	.62	.26	C	42.43	110.00	13.56
4	0	24	6.2		-.38	-.33	-.05	-.75	-.65	-.11		29.13	81.92	12.50
TOTAL		390												

ITEM NUMBER 34

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	4	1.0		-.18	-.11	-.08	-.65	-.39	-.31		27.25	86.25	10.50
2	0	5	1.3		-.21	-.21	-.09	-.70	-.73	-.31		26.60	71.20	10.60
3	0	7	1.8		-.21	-.18	-.04	-.63	-.54	-.13		28.57	80.86	12.14
C 4	1	374	95.9	C	.35	.10	.12	.65	.56	.23	C	40.66	106.14	13.26
TOTAL		390												

ITEM NUMBER 35

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	122	31.3		-.37	-.34	-.04	-.48	-.44	-.05		36.02	96.00	13.01
C 2	1	208	53.3	C	.41	.38	.10	.51	.47	.13	C	42.99	111.41	13.49
3	0	34	8.7		-.06	-.05	-.09	-.11	-.10	-.16		38.68	101.91	12.24
4	0	26	6.7		-.06	-.07	-.03	-.12	-.13	-.06		38.35	100.58	12.77
TOTAL		390												

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 39

TEST NO 1 TOEFL SUBTEST 3 SECTION 3: READING

ITEM NUMBER 36

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	54	13.8		.05	.09	.04	.07	.14	.06		40.96	109.20	13.48
C 2	1	300	76.9	C	.18	.12	.05	.25	.17	.07	C	40.87	106.24	13.27
3	0	31	7.9		-.26	-.24	-.07	-.48	-.44	-.14		33.42	90.13	12.35
4	0	5	1.3		-.19	-.15	-.12	-.64	-.52	-.40		27.80	80.60	9.80
TOTAL		390												

ITEM NUMBER 37

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST MEANS	TT MEANS	EC MEANS
1	0	2	.5		.01	-.02	-.06	.06	-.08	-.29		41.50	101.00	10.50
2	0	9	2.3		-.28	-.24	-.07	-.77	-.66	-.20		26.56	77.00	11.67
3	0	5	1.3		-.20	-.17	-.15	-.67	-.57	-.50		27.20	78.60	9.00
C 4	1	374	95.9	C	.32	.28	.16	.60	.53	.30	C	40.61	106.09	13.29
TOTAL		390												

ITEM NUMBER 38 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	14	3.6		-.23	-.21	-.14	-.53	-.50	-.33		31.36	85.36	10.79
2	0	10	2.6		-.17	-.18	-.09	-.46	-.49	-.25		32.20	84.50	11.30
3	0	45	11.5		-.32	-.29	-.11	-.53	-.47	-.19		33.40	90.62	12.16
C 4	1	321	82.3	C	.45	.42	.20	.66	.61	.29	C	41.69	108.56	13.49
TOTAL		390												

ITEM NUMBER 39 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
C 1	1	212	54.4	C	.37	.30	.12	.46	.38	.15	C	42.65	109.99	13.54
2	0	22	5.6		-.21	-.21	-.07	-.43	-.42	-.14		33.64	89.86	12.27
3	0	43	11.0		-.22	-.17	-.16	-.36	-.28	-.27		35.49	96.40	11.70
4	0	112	28.7		-.14	-.10	.02	-.18	-.13	.03		38.52	102.31	13.28
OTHER	0	1	.3		-.10	-.11	-.05	-.61	-.70	-.32		26.00	66.00	10.00
TOTAL		390												

ITEM NUMBER 40 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	5	1.3		-.16	-.14	.00	-.54	-.48	.00		29.80	82.80	13.20
C 2	1	342	87.7	C	.52	.48	.19	.82	.75	.29	C	41.59	108.32	13.41
3	0	28	7.2		-.35	-.35	-.17	-.66	-.67	-.33		30.79	82.07	11.14
4	0	14	3.6		-.32	-.27	-.07	-.75	-.63	-.17		27.86	80.07	11.93
OTHER	0	1	.3		-.12	-.05	-.05	-.74	-.34	-.32		23.00	86.00	10.00
TOTAL		390												

TEST NO 1 TOEFL — SUBTEST 3 SECTION 3: READING

ITEM NUMBER 41 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	10	2.6		-.20	-.17	-.09	-.53	-.45	-.23		30.90	86.10	11.40
C 2	1	334	85.6	C	.43	.39	.17	.65	.59	.26	C	41.44	107.93	13.41
3	0	32	8.2		-.23	-.23	-.05	-.42	-.42	-.10		34.38	90.88	12.59
4	0	14	3.6		-.30	-.25	-.17	-.71	-.58	-.39		28.43	82.00	10.36
TOTAL		390												

ITEM NUMBER 42 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	18	4.6		-.16	-.14	-.11	-.35	-.31	-.24		34.61	93.39	11.56
2	0	79	20.3		-.20	-.21	-.07	-.29	-.30	-.10		37.13	97.42	12.73
3	0	95	24.4		-.23	-.18	-.13	-.31	-.24	-.18		37.14	99.35	12.44
C 4	1	198	50.8	C	.43	.38	.21	.53	.48	.26	C	43.25	111.87	13.86
TOTAL		390												

ITEM NUMBER 43 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	17	4.4		-.24	-.17	-.11	-.52	-.38	-.25		31.88	90.47	11.47
C 2	1	339	86.9	C	.45	.41	.18	.69	.64	.28	C	41.43	107.94	13.41
3	0	10	2.6		-.26	-.24	-.12	-.68	-.63	-.33		28.30	78.40	10.70
4	0	24	6.2		-.26	-.28	-.08	-.52	-.55	-.16		32.42	85.46	12.17
TOTAL		390												

ITEM NUMBER 44 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
C 1	1	291	74.6	C	.51	.44	.20	.69	.60	.27	C	42.34	109.71	13.56
2	0	23	5.9		-.19	-.16	-.05	-.38	-.33	-.11		34.52	93.17	12.00
3	0	47	12.1		-.34	-.28	-.13	-.55	-.46	-.22		33.32	91.23	12.48
4	0	28	7.2		-.26	-.24	-.12	-.48	-.45	-.22		33.25	89.43	11.82
OTHER	0	1	.3		-.03	.00	-.03	-.18	-.02	-.22		36.00	104.00	11.00
TOTAL		390												

ITEM NUMBER 45 — COEFFICIENTS OF CORRELATION — MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	14	3.6		-.35	-.31	-.12	-.83	-.72	-.29		26.50	76.43	11.07
2	0	14	3.6		-.19	-.14	-.07	-.46	-.32	-.17		32.64	92.14	11.93
3	0	15	3.8		-.21	-.21	-.15	-.48	-.48	-.36		32.33	86.07	10.67
C 4	1	346	88.7	C	.45	.39	.22	.72	.62	.34	C	41.33	107.54	13.43
OTHER	0	1	.3		-.03	.00	-.03	-.18	-.02	-.22		36.00	104.00	11.00

TOTAL 390

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 41

TEST NO 1 TOEFL SUBTEST 3 SECTION 3: READING

ITEM NUMBER 46 COEFFICIENTS OF CORRELATION MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	11	2.8		-.24	-.18	-.12	-.62	-.46	-.30		29.45	86.09	10.91
2	0	14	3.6		-.36	-.30	-.12	-.84	-.71	-.29		26.29	76.86	11.07
3	0	13	3.3		-.29	-.25	-.14	-.71	-.60	-.33		28.31	80.85	10.77
C 4	1	350	89.7	C	.55	.44	.24	.89	.71	.38	C	41.51	107.72	13.44
OTHER	0	2	.5		-.10	-.04	-.06	-.49	-.19	-.29		29.50	95.00	10.50
TOTAL		390												

ITEM NUMBER 47 COEFFICIENTS OF CORRELATION MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
C 1	1	325	83.3	C	.46	.40	.25	.67	.59	.37	C	41.65	108.30	13.54
2	0	12	3.1		-.25	-.20	-.15	-.62	-.50	-.38		29.67	84.75	10.42
3	0	25	6.4		-.29	-.26	-.19	-.56	-.51	-.37		31.96	86.88	10.84
4	0	27	6.9		-.22	-.20	-.08	-.42	-.38	-.15		34.07	91.70	12.26
OTHER	0	1	.3		-.03	.00	-.03	-.18	-.02	-.22		36.00	104.00	11.00
TOTAL		390												

ITEM NUMBER 48 COEFFICIENTS OF CORRELATION MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	11	2.8		-.16	-.18	-.05	-.42	-.46	-.12		33.00	86.18	12.27
C 2	1	364	93.3	C	.24	.21	.09	.42	.36	.16	C	40.60	106.03	13.26
3	0	7	1.8		-.12	-.08	-.03	-.35	-.24	-.09		33.71	94.29	12.43
4	0	7	1.8		-.12	-.08	-.07	-.36	-.25	-.22		33.57	94.00	11.43
OTHER	0	1	.3		-.03	.00	-.03	-.18	-.02	-.22		36.00	104.00	11.00
TOTAL		390												

ITEM NUMBER 49 COEFFICIENTS OF CORRELATION MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	36	9.2		-.23	-.21	.00	-.41	-.37	.01		34.67	93.08	13.22
2	0	15	3.8		-.22	-.17	-.12	-.51	-.39	-.28		31.80	89.93	11.20
C 3	1	299	76.7	C	.45	.40	.11	.62	.55	.16	C	41.98	109.01	13.38
4	0	39	10.0		-.26	-.25	-.08	-.45	-.43	-.14		34.21	91.49	12.38
OTHER	0	1	.3		-.03	.00	-.03	-.18	-.02	-.22		36.00	104.00	11.00
TOTAL		390												

1 LERTAP 2.0 SUMMARY ITEM STATISTICS
PAGE 42

TEST NO 1 TOEFL SUBTEST 3 SECTION 3: READING

ITEM NUMBER 50 COEFFICIENTS OF CORRELATION MEANS

OPTION	WT	N	P		PB-ST	PB-TT	PB-EC	B-ST	B-TT	B-EC		ST	TT	EC
1	0	2	.5		-.14	-.12	-.14	-.68	-.57	-.66		25.50	75.50	7.00
2	0	23	5.9		-.23	-.19	-.09	-.45	-.38	-.17		33.39	91.17	12.04
3	0	14	3.6		-.23	-.19	.02	-.55	-.46	.05		31.14	86.86	13.57
C 4	1	351	90.0	C	.35	.30	.09	.58	.49	.14	C	41.01	106.84	13.27
TOTAL		390												

1 LERTAP 2.0 SUBTEST STATISTICS
PAGE 43

TEST NO 1 TOEFL SUBTEST 3 SECTION 3: READING

NUMBER OF INDIVIDUALS	= 390.00	NUMBER OF ITEMS	= 50.00
MEAN	= 40.12	HIGHEST SCORE	= 50.00
STANDARD DEVIATION	= 7.48	LOWEST SCORE	= 11.00

SOURCE OF VARIANCE	D.F.	S.S.	M.S.
INDIVIDUALS	389.00	435.64	1.12
ITEMS	49.00	330.26	6.74
RESIDUAL	19061.00	2325.18	.12

TOTAL	19499.00	3091.08	.16

HOYT ESTIMATE OF RELIABILITY = .89

STANDARD ERROR OF MEASUREMENT = 2.44

1 LERTAP 2.0 TOTAL TEST STATISTICS
PAGE 44

TEST NO 1 TOEFL

NUMBER OF INDIVIDUALS	= 390.00	NUMBER OF ITEMS	= 140.00
MEAN	= 105.04	HIGHEST SCORE	= 134.00
STANDARD DEVIATION	= 18.11	LOWEST SCORE	= 50.00

SOURCE OF VARIANCE	D.F.	S.S.	M.S.
INDIVIDUALS	389.00	911.66	2.34
ITEMS	139.00	1567.93	11.28
RESIDUAL	54071.00	7750.41	.14
TOTAL	54599.00	10230.00	.19

HOYT ESTIMATE OF RELIABILITY = .94

STANDARD ERROR OF MEASUREMENT = 4.46

NO. SUBTESTS WITH NON-ZERO WT = 3.00

CRONBACHS ALPHA FOR COMPOSITE = .84

1 LERTAP 2.0 HISTOGRAM
PAGE 45

TEST NO 1 TOEFL SUBTEST 1 SECTION 1: LISTENING

NUMBER OF OBSERVATIONS = 390 MEAN = 33.48 S.D. = 8.06 LOWEST SCORE = 11.00, HIGHEST SCORE= 49.00

Z	LB	UB	F	P	CF	CP	EACH * REPRESENTS 1 OBSERVATIONS
-2.79	10.50	11.50	1	.26	1	.26	*
-2.66	11.50	12.50	0	.00	1	.26	
-2.54	12.50	13.50	0	.00	1	.26	
-2.42	13.50	14.50	1	.26	2	.51	*
-2.29	14.50	15.50	2	.51	4	1.03	**
-2.17	15.50	16.50	2	.51	6	1.54	**
-2.04	16.50	17.50	6	1.54	12	3.08	******
-1.92	17.50	18.50	5	1.28	17	4.36	*****
-1.80	18.50	19.50	4	1.03	21	5.38	****
-1.67	19.50	20.50	6	1.54	27	6.92	******
-1.55	20.50	21.50	9	2.31	36	9.23	*********
-1.42	21.50	22.50	6	1.54	42	10.77	******
-1.30	22.50	23.50	8	2.05	50	12.82	********
-1.18	23.50	24.50	10	2.56	60	15.38	**********
-1.05	24.50	25.50	16	4.10	76	19.49	****************
-.93	25.50	26.50	13	3.33	89	22.82	*************
-.80	26.50	27.50	9	2.31	98	25.13	*********
-.68	27.50	28.50	8	2.05	106	27.18	********
-.56	28.50	29.50	13	3.33	119	30.51	*************
-.43	29.50	30.50	11	2.82	130	33.33	**********
-.31	30.50	31.50	14	3.59	144	36.92	**************
-.18	31.50	32.50	19	4.87	163	41.79	*******************
-.06	32.50	33.50	17	4.36	180	46.15	*****************
.06	33.50	34.50	24	6.15	204	52.31	************************
.19	34.50	35.50	16	4.10	220	56.41	****************
.31	35.50	36.50	19	4.87	239	61.28	*******************
.44	36.50	37.50	17	4.36	256	65.64	*****************
.56	37.50	38.50	17	4.36	273	70.00	*****************
.69	38.50	39.50	16	4.10	289	74.10	****************
.81	39.50	40.50	18	4.62	307	78.72	******************
.93	40.50	41.50	14	3.59	321	82.31	**************
1.06	41.50	42.50	13	3.33	334	85.64	*************
1.18	42.50	43.50	11	2.82	345	88.46	**********
1.31	43.50	44.50	17	4.36	362	92.82	*****************
1.43	44.50	45.50	8	2.05	370	94.87	********
1.55	45.50	46.50	8	2.05	378	96.92	********
1.68	46.50	47.50	5	1.28	383	98.21	*****
1.80	47.50	48.50	5	1.28	388	99.49	*****
1.93	48.50	49.50	2	.51	390	100.00	**
2.05	49.50	50.50	0	.00	390	100.00	

1 LERTAP 2.0 HISTOGRAM
PAGE 46

TEST NO 1 TOEFL SUBTEST 2 SECTION 2: STRUCTURE

NUMBER OF OBSERVATIONS = 390 MEAN = 31.44 S.D. = 4.85 LOWEST SCORE = 9.00, HIGHEST SCORE= 40.00

Z	LB	UB	F	P	CP	CP	EACH * REPRESENTS 1 OBSERVATIONS
-5.45	4.50	5.50	0	.00	0	.00	
-5.25	5.50	6.50	0	.00	0	.00	
-5.04	6.50	7.50	0	.00	0	.00	
-4.83	7.50	8.50	0	.00	0	.00	
-4.63	8.50	9.50	1	.26	1	.26	*
-4.42	9.50	10.50	1	.26	2	.51	*
-4.21	10.50	11.50	0	.00	2	.51	
-4.01	11.50	12.50	0	.00	2	.51	
-3.80	12.50	13.50	0	.00	2	.51	
-3.60	13.50	14.50	0	.00	2	.51	
-3.39	14.50	15.50	1	.26	3	.77	*
-3.18	15.50	16.50	0	.00	3	.77	
-2.98	16.50	17.50	0	.00	3	.77	
-2.77	17.50	18.50	5	1.28	8	2.05	*****
-2.56	18.50	19.50	4	1.03	12	3.08	****
-2.36	19.50	20.50	3	.77	15	3.85	***
-2.15	20.50	21.50	2	.51	17	4.36	**
-1.95	21.50	22.50	2	.51	19	4.87	**
-1.74	22.50	23.50	7	1.79	26	6.67	*******
-1.53	23.50	24.50	6	1.54	32	8.21	******
-1.33	24.50	25.50	10	2.56	42	10.77	**********
-1.12	25.50	26.50	11	2.82	53	13.59	***********
-.92	26.50	27.50	16	4.10	69	17.69	****************
-.71	27.50	28.50	18	4.62	87	22.31	******************
-.50	28.50	29.50	23	5.90	110	28.21	************************
-.30	29.50	30.50	25	6.41	135	34.62	*************************
-.09	30.50	31.50	34	8.72	169	43.33	**********************************
.12	31.50	32.50	46	11.79	215	55.13	**
.32	32.50	33.50	32	8.21	247	63.33	********************************
.53	33.50	34.50	28	7.18	275	70.51	****************************
.73	34.50	35.50	38	9.74	313	80.26	**************************************
.94	35.50	36.50	25	6.41	338	86.67	*************************
1.15	36.50	37.50	29	7.44	367	94.10	*****************************
1.35	37.50	38.50	15	3.85	382	97.95	***************
1.56	38.50	39.50	6	1.54	388	99.49	******
1.77	39.50	40.50	2	.51	390	100.00	**
1.97	40.50	41.50	0	.00	390	100.00	
2.18	41.50	42.50	0	.00	390	100.00	
2.38	42.50	43.50	0	.00	390	100.00	
2.59	43.50	44.50	0	.00	390	100.00	

1 LERTAP 2.0 HISTOGRAM
PAGE 47

TEST NO 1 TOEFL SUBTEST 3 SECTION 3: READING

NUMBER OF OBSERVATIONS = 390 MEAN = 40.12 S.D. = 7.48 LOWEST SCORE = 11.00, HIGHEST SCORE= 50.00

Z	LB	UB	F	P	CF	CP	EACH * REPRESENTS 1 OBSERVATIONS
-3.89	10.50	11.50	1	.26	1	.26	*
-3.76	11.50	12.50	1	.26	2	.51	*
-3.62	12.50	13.50	0	.00	2	.51	
-3.49	13.50	14.50	0	.00	2	.51	
-3.36	14.50	15.50	0	.00	2	.51	
-3.22	15.50	16.50	0	.00	2	.51	
-3.09	16.50	17.50	0	.00	2	.51	
-2.96	17.50	18.50	0	.00	2	.51	
-2.82	18.50	19.50	0	.00	2	.51	
-2.69	19.50	20.50	1	.26	3	.77	*
-2.56	20.50	21.50	5	1.28	8	2.05	*****
-2.42	21.50	22.50	3	.77	11	2.82	***
-2.29	22.50	23.50	4	1.03	15	3.85	****
-2.15	23.50	24.50	3	.77	18	4.62	***
-2.02	24.50	25.50	7	1.79	25	6.41	*******
-1.89	25.50	26.50	4	1.03	29	7.44	****
-1.75	26.50	27.50	4	1.03	33	8.46	****
-1.62	27.50	28.50	5	1.28	38	9.74	*****
-1.49	28.50	29.50	3	.77	41	10.51	***
-1.35	29.50	30.50	4	1.03	45	11.54	****
-1.22	30.50	31.50	9	2.31	54	13.85	*********
-1.09	31.50	32.50	9	2.31	63	16.15	*********
-.95	32.50	33.50	7	1.79	70	17.95	*******
-.82	33.50	34.50	11	2.82	81	20.77	**********
-.68	34.50	35.50	13	3.33	94	24.10	**********
-.55	35.50	36.50	12	3.08	106	27.18	**********
-.42	36.50	37.50	16	4.10	122	31.28	**********
-.28	37.50	38.50	9	2.31	131	33.59	*********
-.15	38.50	39.50	13	3.33	144	36.92	**********
-.02	39.50	40.50	21	5.38	165	42.31	**********
.12	40.50	41.50	19	4.87	184	47.18	**********
.25	41.50	42.50	22	5.64	206	52.82	**********
.38	42.50	43.50	16	4.10	222	56.92	**********
.52	43.50	44.50	27	6.92	249	63.85	**********
.65	44.50	45.50	30	7.69	279	71.54	**********
.79	45.50	46.50	37	9.49	316	81.03	**********
.92	46.50	47.50	24	6.15	340	87.18	**********
1.05	47.50	48.50	23	5.90	363	93.08	**********
1.19	48.50	49.50	17	4.36	380	97.44	**********
1.32	49.50	50.50	10	2.56	390	100.00	**********

1 LERTAP 2.0 HISTOGRAM
PAGE 48

TEST NO 1 TOEFL TOTAL TEST

NUMBER OF OBSERVATIONS = 390 MEAN = 105.04 S.D. = 18.11 LOWEST SCORE = 50.00, HIGHEST SCORE= 134.00

Z	LB	UB	F	P	CF	CP	EACH * REPRESENTS 1 OBSERVATIONS
-3.04	48.92	51.08	1	.26	1	.26	*
-2.92	51.08	53.23	0	.00	1	.26	
-2.80	53.23	55.38	1	.26	2	.51	*
-2.68	55.38	57.54	1	.26	3	.77	*
-2.56	57.54	59.69	0	.00	3	.77	
-2.44	59.69	61.85	1	.26	4	1.03	*
-2.33	61.85	64.00	9	2.31	13	3.33	*********
-2.21	64.00	66.15	5	1.28	18	4.62	*****
-2.09	66.15	68.31	2	.51	20	5.13	**
-1.97	68.31	70.46	2	.51	22	5.64	**
-1.85	70.46	72.62	3	.77	25	6.41	***
-1.73	72.62	74.77	2	.51	27	6.92	**
-1.61	74.77	76.92	6	1.54	33	8.46	******
-1.49	76.92	79.08	10	2.56	43	11.03	**********
-1.37	79.08	81.23	7	1.79	50	12.82	*******
-1.25	81.23	83.38	5	1.28	55	14.10	*****
-1.14	83.38	85.54	8	2.05	63	16.15	********
-1.02	85.54	87.69	7	1.79	70	17.95	*******
-.90	87.69	89.85	9	2.31	79	20.26	*********
-.78	89.85	92.00	8	2.05	87	22.31	********
-.66	92.00	94.15	7	1.79	94	24.10	*******
-.54	94.15	96.31	8	2.05	102	26.15	********
-.42	96.31	98.46	15	3.85	117	30.00	***************
-.30	98.46	100.62	11	2.82	128	32.82	**********
-.18	100.62	102.77	22	5.64	150	38.46	**********
-.07	102.77	104.92	18	4.62	168	43.08	**********
.05	104.92	107.08	27	6.92	195	50.00	**********
.17	107.08	109.23	23	5.90	218	55.90	**********
.29	109.23	111.38	17	4.36	235	60.26	**********
.41	111.38	113.54	18	4.62	253	64.87	**********
.53	113.54	115.69	10	2.56	263	67.44	**********
.65	115.69	117.85	17	4.36	280	71.79	**********
.77	117.85	120.00	24	6.15	304	77.95	**********
.89	120.00	122.15	11	2.82	315	80.77	**********
1.00	122.15	124.31	15	3.85	330	84.62	***************
1.12	124.31	126.46	17	4.36	347	88.97	**********
1.24	126.46	128.62	17	4.36	364	93.33	**********
1.36	128.62	130.77	11	2.82	375	96.15	**********
1.48	130.77	132.92	11	2.82	386	98.97	**********
1.60	132.92	135.08	4	1.03	390	100.00	****

```
1     LERTAP 2.0                                        HISTOGRAM
PAGE  49

      TEST NO  1   TOEFL                                          EXTERNAL CRITERION


      NUMBER OF OBSERVATIONS =  390     MEAN =   13.18     S.D. =    3.27     LOWEST SCORE =   4.00,   HIGHEST SCORE=
34.00



          Z       LB       UB       F      P       CF     CP    EACH * REPRESENTS  1 OBSERVATIONS

        -4.04     -.50      .50      0     .00      0     .00
        -3.73      .50     1.50      0     .00      0     .00
        -3.42     1.50     2.50      0     .00      0     .00
        -3.12     2.50     3.50      0     .00      0     .00
        -2.81     3.50     4.50      2     .51      2     .51  **
        -2.50     4.50     5.50      0     .00      2     .51
        -2.20     5.50     6.50      3     .77      5    1.28  ***
        -1.89     6.50     7.50      7    1.79     12    3.08  *******
        -1.59     7.50     8.50      5    1.28     17    4.36  *****
        -1.28     8.50     9.50     31    7.95     48   12.31  *******************************
         -.97     9.50    10.50     42   10.77     90   23.08  ******************************************
         -.67    10.50    11.50     30    7.69    120   30.77  ****************************
         -.36    11.50    12.50     49   12.56    169   43.33  ************************************************
         -.05    12.50    13.50     42   10.77    211   54.10  ******************************************
          .25    13.50    14.50     40   10.26    251   64.36  ****************************************
          .56    14.50    15.50     49   12.56    300   76.92  ************************************************
          .86    15.50    16.50     32    8.21    332   85.13  ********************************
         1.17    16.50    17.50     29    7.44    361   92.56  *****************************
         1.48    17.50    18.50     16    4.10    377   96.67  ****************
         1.78    18.50    19.50      6    1.54    383   98.21  ******
         2.09    19.50    20.50      2     .51    385   98.72  **
         2.39    20.50    21.50      2     .51    387   99.23  **
         2.70    21.50    22.50      2     .51    389   99.74  **
         3.01    22.50    23.50      0     .00    389   99.74
         3.31    23.50    24.50      0     .00    389   99.74
         3.62    24.50    25.50      0     .00    389   99.74
         3.93    25.50    26.50      0     .00    389   99.74
         4.23    26.50    27.50      0     .00    389   99.74
         4.54    27.50    28.50      0     .00    389   99.74
         4.84    28.50    29.50      0     .00    389   99.74
         5.15    29.50    30.50      0     .00    389   99.74
         5.46    30.50    31.50      0     .00    389   99.74
         5.76    31.50    32.50      0     .00    389   99.74
         6.07    32.50    33.50      0     .00    389   99.74
         6.37    33.50    34.50      1     .26    390  100.00  *
         6.68    34.50    35.50      0     .00    390  100.00
         6.99    35.50    36.50      0     .00    390  100.00
         7.29    36.50    37.50      0     .00    390  100.00
         7.60    37.50    38.50      0     .00    390  100.00
         7.91    38.50    39.50      0     .00    390  100.00
1     LERTAP 2.0                                        CORRELATIONS
PAGE  50

      TEST NO  1   TOEFL

         VARIABLE      TYPE            NAME

             1       SUBTEST  1   SECTION 1: LISTENING

             2       SUBTEST  2   SECTION 2: STRUCTURE

             3       SUBTEST  3   SECTION 3: READING

             4       TOTAL TEST

             5       EXT. CRIT.
1     LERTAP 2.0                                        CORRELATIONS
PAGE  51

       VAR      1        2        3        4        5

         1   1.000     .661     .686     .905     .322

         2    .661    1.000     .677     .842     .374

         3    .686     .677    1.000     .900     .386

         4    .905     .842     .900    1.000     .403

         5    .322     .374     .386     .403    1.000
1
```