TEST-TAKING STRATEGY USE ON THE READING SECTION OF THE TOEFL IBT: A STUDY OF ARAB ESL LEARNERS

By

MOHAMMED S ASSIRI

Bachelor of Arts in English
King Khalid University
Abha, Saudi Arabia
2000

Master of Arts in Teaching English as
a Second Language
University of Kansas
Lawrence, Kansas
2004

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2011

TEST-TAKING STRATEGY USE ON THE READING SECTION OF THE TOEFL IBT: A STUDY OF ARAB ESL LEARNERS

Dissertation Approved:

Dr. Gene Halleck

Dissertation Adviser

Dr. Ravi Sheorey

Dr. Susan Garzon

Dr. Nadine Olson

Outside Committee Member

Dr. Mark E. Payton

Dean of the Graduate College

ACKNOWLEDGMENTS

A dissertation journey cannot make its way to its destination without involvement of many people. I owe my profound gratitude and admiration to my advisor and chair of my dissertation committee, Dr. Gene Halleck. Throughout my doctoral study and research, Dr. Halleck's care, guidance, and support have always been up to my needs, if not beyond.

I would like to extend my sincere tribute and appreciation to the members of my advisory and dissertation committee: Dr. Ravi Sheorey, Dr. Susan Garzon, and Dr. Nadine Olson. I have always found their advice invaluable over and above their indispensable feedback. I highly esteem Dr. Olson's commitment to serve on my committee despite her circumstances.

To my parents, I dedicate all my successes to you for you are the reason for my being who I am. You have kept encouraging and praying for me all the way. To Fatimah, Abdullah, and Shuaib—my small family, I would not have been as diligent in my doctoral work without having you beside me. To my siblings and relatives, I can only say thank you for your concern and interest.

Best regards to my respondents whose participation in my research could not be more appreciated. Special thanks to my friends who have always been cheering and reassuring.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Background of the Study

'The TOEFL is a test of strategies!' to quote one of my Libyan acquaintances who has been struggling for more than a year to achieve a high score on the Test of English as a Foreign Language (henceforth, TOEFL). From brief talks about the TOEFL with a number of Arab students who are attending a south-western university in the United States of America, it appeared to me that most of them, if not all, are aware of the import of developing and practicing strategies that aid their performance on the TOEFL, especially on the reading section. This is due in part to the difficulty of responding to the question items on this section of the test by merely understanding the reading texts, which in turn motivates test takers to save their time by using strategies that can help them achieve high scores without having to exert strenuous effort to understand the text. (It is worth bringing to the reader's attention that the acronym TOEFL is used without an ending to refer to the test in general or its framework. However, wherever there is a need to refer to a specific version of the test (e.g., TOEFL-iBT), an ending is attached to the acronym TOEFL to reflect the given version).

The researcher has had the chance to meet with a number of Arab ESL students who were in the midst of preparation to take the TOEFL. Regardless of the test version they were planning to take, be it paper-based (TOEFL-PBT) or internet-based (TOEFL iBT), these folks spend weeks, if not months, responding to as many practice tests as they can. Principally, they see this form of preparation as 'practice' that helps them develop the kinds of strategies that could increase their chance of getting as many question items correct as possible on the actual test. Certainly, this form of practice is not expected to make a noticeable difference in their language and reading proficiency, but at least enables them to think of ways of how to deal with question items on the official TOEFL. Usually, the preparation to take the TOEFL among these folks is more of a collaborative activity, which allows them to share their experiences with the test and strategies they believe to be helpful. Many of them think that enrolling in a program of intensive English may help build up their vocabulary and structure repertoire and sharpen their language skills, but does not equip them with strategies and tips that can help them pass the TOEFL. By far, these folks believe that being acquainted with such strategies and tips as those applicable to the TOEFL is a necessity because when responding to a reading set that is part of the reading section of the test, for example, 'one cannot read the whole passage once and answer all question items based on one reading of the passage'.

Given the above case in point, it becomes evident that the kinds of strategies test takers resort to in order to manage their response or compensate for their lack of ability to achieve full understanding of the text on tests of reading comprehension (henceforth, test-taking strategies) are worth researching and exploring. And, because the version of the TOEFL that is most frequent in use at present is the internet-based (henceforth, the

TOEFL iBT), it would be interesting to see how test-taking strategy use on the TOEFL-iBT reading section interacts with other factors that are part of the test-taking process, for example, test tasks and formats, language and reading proficiency, test performance, cultural and linguistic backgrounds, and cognitive processing. More importantly, the study of test-taking strategies in relation to these factors and others can enlighten us about the extent to which performance on language tests, including those of reading comprehension, can be noticeably shaped by factors other than language proficiency (Yoshizawa, 2002). This is especially the case with such a high-stakes test as the TOEFL iBT whose scores make a difference in academic and professional lives of millions of people around the globe.

Strategy use on the reading section of the TOEFL-iBT: Localizing research efforts.

The test-taking process represents a problem-solving activity (Smith, 1979), calling for the use of strategies or tactics that can help deal with this process successfully. And, since members of diverse cultural domains differ in how they approach and deal with problems, examinees who belong to these cultures may exhibit varied levels of performance on language tests (Duran, 1989). As a result, one would expect that such cross-culturally varied perceptions may be more likely to influence the ways in which test takers employ strategies and perform language tasks. In view of that, the study of test-taking strategy use among a population of examinees who belong to specific cultural and linguistic backgrounds can potentially prove useful in drawing inferences about strategy use applicable to this population.

Scholars in the field of language testing and assessment (e.g., Phakiti, 2008; Rupp, Ferne, & Choi, 2006) have expressed the need for more research that targets specific testing contexts in terms of both the testing instrument and the population of examinees. Phakiti (2008) justified his position by considering that both language competence and performance are variable across different settings, and so contingent on aspects that typify the specific setting where they are researched. On the other hand, Rupp et al. (2006) were more interested in the nature of the relationships that link such variables as the technical aspects of a given test, test-taking strategy use, and characteristics of test takers. According to Rupp and his colleagues (2006), this can be made possible by means of focus groups. The members of each group share some common characteristics (e.g., linguistic and/or cultural background).

Owing to both scarce research and broad generalizations, the field of language testing and assessment still lacks a theory that encompasses what we have come to know about strategy use on language tests (Cohen, 2006). Language testing specialists are yet to settle on the directionality of the relationship that joins together language proficiency, task specifications, and performance (Cohen, 2006). Undue attention used to be paid to the products of the test that take the form of numerical values detached from explanatory information on how these three factors relate to one another (N. Anderson, 1991). Only recently has this area of inquiry been approached by focusing on the processes involved in test-taking. For example, the study of strategy use on tests of reading comprehension can reveal much more information on how testees arrived at their answers than do numerical or symbolic estimates of performance (N. Anderson, 1991).

Thus far, research within the field of language testing and assessment has placed more emphasis on strategy use on general reading tasks (e.g., cloze, multiple-choice, open-ended), rather than on strategies on tasks that are part of specific tests such as high-stakes ones (cf., Tian, 2000). There have been a few studies that looked into strategy use on the reading subtest of the TOEFL (e.g., Cohen & Upton, 2006; Tian, 2000). Cohen and Upton's (2006) study was the only attempt made to relate strategy use to task and item types on the reading subtest of the TOEFL iBT. In their study, however, there is a considerable overlap between strategy descriptions across several task and item types, and a few strategies appeared to be specific to certain task and item types. This may point out prevalent instances of inaccurate coding of strategies, which made it difficult to give more specific descriptions to individual strategies in relation to task and item types.

Now, with technological advancements in the test-making industry, there is an ardent demand for ongoing research that parallels the revolution of high-stakes language tests such as the TOEFL, the International English Language Testing System (or the IELTS), and others. Certainly, if such research is realized, it will: first, inform test makers' efforts to evaluate and enhance language tests; second, respond to test takers' need for guidance to the best ways of preparing for these tests; and third, enhance the quality of language teaching programs through positive washback. Through the study of proficiency-related aspects of strategy use, we can also devise practical implications for test preparation programs by placing more emphasis on these aspects, especially with low-ability language learners (N. Anderson, 2005). Such aspects make up the core of another pertinent area of research, to be foregrounded next.

Proficiency level and strategy use: How do the two relate on the TOEFL-iBT reading section?

While research on strategy use on language learning tasks points to notable discrepancies between high- and low- proficiency learners, questions have been raised as to how these discrepancies among different levels of proficiency manifest themselves on language tests. It is assumed that since proficiency tests were designed with the intention of distinguishing levels of proficiency one from another, response behaviors and strategies would expectedly vary among test takers at different levels of proficiency. The fact that variation in proficiency level explains variability in performance on language tests may be accounted for in terms of response strategies and behaviors. Hence, the existence of certain strategies or behaviors in contrast to others may characterize high performance, whereas their absence may be typical of low performance.

The findings of previous research in regard to differential use of strategies by proficiency levels on language tests (e.g., C. Gordon, 1987; Mangubhai, 1990, 1991; N. Anderson, 1991; Nevo, 1989; Nikolov, 2006; Phakiti, 2003) remain general in their foci, and so are not applicable to specific contexts such as the TOEFL-iBT. On the basis of their study of test-taking strategy use on the TOEFL-iBT reading subtest, Cohen and Upton (2006) surmised that their high-proficiency respondents used certain strategies in their responses to the reading tasks they would not have used had they been at a lower level of proficiency. This observation can certainly give rise to the following research questions, among others, in regard to strategy use on the TOEFL-iBT reading tasks: How would high- and low-proficiency test takers perform on the reading tasks? What strategies would each proficiency group be more disposed to use compared with the other

one? And, how would the strategies used by each group fit on the scale of effective strategy use?

A number of studies have sought to investigate how high- and low-proficiency test takers perform on reading tasks and how they differ from each other in this respect (e.g., C. Gordon , 1987; Mangubhai, 1990, 1991; N. Anderson, 1991; Nevo, 1989; Nikolov, 2006; Phakiti, 2003; Tian, 2000; Yamashita, 2003; Yoshizawa, 2002). Taken together, these studies concluded that test takers with discrete levels of proficiency tend to approach reading tasks differently from each other and exhibit different response behaviors and strategies. Two of these studies aimed at earlier versions of the TOEFL; that is, where Tian (2000) examined strategy use by proficiency levels on the TOEFL-PBT, Yoshizawa (2002) used the computer-based version of the test (TOEFL-CBT) for almost the same purpose. This is to say that differential use of test-taking strategies by proficiency levels remains an unexplored research venue in the context of the TOEFL-iBT reading section.

Researchers (e.g., C. Gordon , 1987; Mangubhai, 1990, 1991; Phakiti, 2003) have noted that one of the most salient differences between high- and low-proficiency test takers is the rate at which each group can process language information. That is, the advanced ability of the former group enables them to benefit more from automatic processing than does the latter. In fact, the latter group may show signs of strategic processing which, owing to its more deliberate nature, is slower than automatic processing. The difference in the rate of language processing can largely determine both the quantity and quality of strategies to be used on language tasks. For example, an

advanced-ability test taker may answer a given test question without having to use as many or as test-wise strategies as those used by a low-ability test taker.

Another line of research has addressed a proficiency-related use of test-management versus the use of test-wiseness strategies. Several studies (e.g., Allan, 1992; Cohen & Upton, 2006; Yoshizawa, 2002) suggested that test takers may try to make use of their problem-solving skills in ways contrary to the purposes for which the test was designed. Such tendencies on the part of test takers are referred to as test-wiseness strategies. According to the aforementioned studies, the use of test-management versus test-wiseness strategies has to do with proficiency level such that the higher the proficiency level of a test taker, the more the likelihood that she will make more use of test-management strategies than test-wiseness strategies. Consequently, it would be interesting to see if the difference between proficiency levels in regard to the use of test-management versus test-wiseness applies to the case of test takers responding to the TOEFL-iBT reading section. Further, it would also be interesting to find out if test takers who have had some experience with the TOEFL-iBT reading tasks would still trust test-wiseness strategies if the need to use those strategies arises.

Some of the previous studies of test-taking strategy use on reading comprehension tests looked for differences between proficiency levels in terms of efficacy of strategy use. For example, research by N. Anderson (1991), Nevo (1989), and Nikolov (2006) has shown that high-proficiency test takers make more effective strategy use compared with their low-proficiency peers. Aspects of effective strategy use include awareness of what strategies to use, when and how to use these strategies, and how to orchestrate strategies with one another (N. Anderson, 1991). Other aspects that can influence efficacy of

strategy use include compatibility of strategies with the test format and task types (Nikolov, 2006). The fact that there are three reading task types on the TOEFL iBT and within these tasks are item types that differ among themselves, as will be shown later, makes the study of efficacy of strategy use by proficiency levels on these tasks and items a valuable venture.

On tests of reading comprehension, it is assumed that proficiency correlates with test performance which, in turns, correlates with the test score. This suggests that in our exploration of differences between proficiency levels in their test-taking strategy use, we can use their scores to distinguish high- from low- proficiency test takers. This necessitates that we use tests of reading proficiency that have satisfactory indices of reliability. As stated above, proficiency level manifests itself in the domain of effective strategy use which, in turn, is expected to contribute to test performance. However, can low-proficiency/scoring test takers utilize aspects of strategy efficacy other than those that are proficiency-related? Another area of inquiry into test-taking strategy use on reading comprehension tests revolves about this question, which speaks to aspects of effective strategy use as determined by test performance.

Test-taking strategies and test performance: Aspects of effective strategy use on the TOEFL-iBT reading section.

Until the beginning of the 1980s, agencies concerned with the development of language tests had lent minimal support to the role of strategy training and instruction in maximizing test takers' chance of performing well and achieving high scores on tests administered by these agencies (Tian, 2000). However, subsequent research (e.g., Powers, 1993; Wadden & Hilke, 1999) has proven that training and instruction of test-

taking strategies could bring about a noticeable improvement in test performance, and so attainment of higher scores. This points to the practical value of providing prospective test takers with training in using test-taking strategies with the aim of helping them perform well and score high on language tests.

According to the Bachman and Palmer (1996) model of communicative language ability, strategy use mediates between competence and performance, and so it influences how competence contributes to performance. In theory, test-taking strategies are like their learning counterparts in that if they are used competently they are certain to bring about positive results. In general, test-taking strategy use on language tests can make a difference in performance (Bachman & Palmer, 1996). Two test takers who are similarly proficient may perform differently on a given language test as a result of each one's using a set of strategies that set him apart from the other (Bachman, 1990). However, drawing on the views of Phakiti (2008) and Rupp et al. (2006), the extent to which this difference affects performance is obviously dependent on the format of the test and the nature of the setting where the test is administered. Hence, one would expect that the kinds of strategies used in response to the reading tasks of the TOEFL iBT and how test takers go about making use of these strategies are both specific to the types of reading tasks on this test.

Scholars concerned with L1 testing research (e.g., Cordón & Day, 1996) have pointed to a need to study the efficacy of test-taking practices and strategies in relation to specific task types and formats. On the other hand, specialists in L2 testing have called for research into how test-taking strategies relate to test performance and whether there are certain strategies that, if used, could result in successful answers to test questions

(e.g., Cohen, 1998; Cohen & Upton, 2006; Yien, 2001). L2 testing researchers (e.g., Nikolov, 2006) have seen the value of this research in identifying and describing how test-taking strategies can be used effectively. To date, there have been few efforts to address the need for this kind of research, and none of these efforts has aimed at the reading section of the TOEFL iBT in particular. Even those studies that have pursued this research (e.g., Nevo, 1989) dealt with reading tasks that were not part of specific tests that involve high-stakes decisions. An important, practical implication of finding out which test-taking strategies are more effective than others as well as ways of how to use strategies efficiently is the possibility of sharing the resulting knowledge of this research with generations of TOEFL-iBT test takers. The merit of such strategic knowledge is that it derives its support from sound, empirical research rather than hunches that may not prove practicable.

The use of a given strategy on a test of reading comprehension can either facilitate or debilitate test performance. Hence, strategization on reading tests calls for careful choice and effective use of strategies on the part of the test taker. This demands strategic awareness of whether or not a given strategy suits the test format and skillfulness in how to manipulate strategy use to fulfill the demands of the test (Nikolov, 2006; Tian, 2000). A strategy used in response to a given question item can be said to be effective so long as its use can result in the provision of the correct answer. However, strategies that have shown to be effective may lose their quality of being effective because they lack some or all of the aspects that render them as such. Such aspects of effective strategy use embrace answers to questions of which strategy to use, when to use it, whether to use it alone or in conjunction with other strategies, how to use strategies both in order and in chorus, and

how to assess and monitor strategy use? (cf., N. Anderson, 1991, 2005). It remains unknown how aspects of effective test-taking strategy use exhibit themselves on the reading tasks of the TOEFL iBT, in spite of the practical value of such an inquiry.

Definition of Key Terms

Among the key terms that will dominate the theory and research presented in this study are the following:

Test-taking Strategies

Test-taking Strategies represent problem-solving techniques or tactics that test takers resort to when trying to answer question items on a given language test or form of assessment (Cohen, 1992, 1998). Test-taking strategies comprise test-management and test-wiseness strategies, the former involving the use of construct-relevant response behaviors and the latter employing textual and/or technical aspects of the test that are not construct-relevant (Allan, 1992; Cohen, 2006; Phakiti, 2008). In this sense, a given test-taking strategy is considered contributory or effective so long as its use results in getting the answer to a given question right; if not, this strategy is noncontributory or ineffective (Nevo, 1989).

Test Format

According to the *Dictionary of language testing* (Davies et al., 2002), test format "refers to the overall design of the test" (p. 200). As such, test format covers such aspects as length of the test, types and numbers of tasks and items, and expected forms of response (Davies et al., 2002). Test format represents essential information that test takers should know before they sit for the test. Research (e.g., Kobayashi, 2002) has indicated that the nature of the test tasks considerably determines test performance as

evidenced in test scores. It has also been confirmed that different test formats tap into distinct reading skills (e.g., Tsagari, 1994).

Test Performance

In Bachman and Palmer's (1996) model of communicative language ability, language ability comprises language knowledge, or "a set of specific knowledge components that are utilized in communication via language" (Bachman, 1990, p. 84) and strategic competence, or "the mental capacity for implementing the components of language competence in contextualized communicative language use" (Bachman, 1990, p. 106). In the context of reading comprehension, language knowledge splits into grammatical knowledge (i.e., of morphology, lexicology, and syntax) and textual knowledge (i.e., of cohesive devices, elements of coherence, and discourse markers) (Bachman, 1990). Strategic competence, on the other hand, serves both compensatory and supplementary manipulation of textual and technical components of information in terms of cognitive and metacognitive strategies in reading comprehension (Rupp, Vock, Harsch, & Köller, 2008) and test-taking strategies in reading test-taking (Davies et al., 2002).

Performance on language tests typically correlates with the level of language ability. It is considerably determined by how test taker characteristics (e.g., background knowledge, cognitive style, anxiety, etc.) and facets of test format act together (Bachman, 1990; Bachman & Palmer, 1996; Douglas, 2002). Hence, test performance in this research was taken to reflect both language knowledge and strategic competence as they apply to the processes of reading comprehension and reading test-taking. Consequently, the use of the TOEFL-iBT reading tasks in this study represented a measure of the

interaction of reading knowledge, facets of test format, and test-taking strategies among participants. The outcomes of such an interaction manifested themselves in the form of the scores the participants received.

TOEFL

The TOEFL refers to the world-wide renowned Test of English as a Foreign Language, administered by the Educational Testing Service (ETS). The TOEFL measures test takers' ability to use academic English for admission purposes to programs of study at the college or university level in English-speaking countries (ETS, 2007). The TOEFL has been administered in three mediums: paper, computer, and internet, each of which has relatively distinct modalities and design features. The TOEFL iBT stands for the internet-based TOEFL, which has been in use since 2005. As its name suggests, this latest version of the TOEFL is administered world-wide via the internet. It has four sections that include listening, speaking, reading, and writing, each of which is assigned a score up to 30, with the total score of the test being 120.

TOEFL-iBT Reading Section

The reading section of the TOEFL iBT is the first section on the test. It aims to assess "the test taker's ability to understand university-level academic texts and passages" (ETS, 2007, p. 8). The reading section of the TOEFL iBT has been intended to imbed three purposes of academic reading, namely reading to find information, reading for basic comprehension, and reading to learn. Accordingly, this section has three reading task types: basic comprehension, inferencing, and reading to learn. On a single administration of the TOEFL iBT, the reading section may include three to five reading sets, each with a 600-700 word text followed by 12 to 14 question items. The total time

allotted for completing this section ranges from 60 to 100 minutes depending on the number of reading sets that appear on the test, at a rate of 20 minutes per reading set.

A reading text can be either expository, which describes and explains a certain topic; argumentative, which discusses a topic from a certain perspective using supporting evidence; or historical, which offers a chronological account of some event (ETS, 2007). The content and level of the text material resemble those of introductory readings assigned at the university level. An examinee expects to deal with a range of topics representing various academic disciplines; however, little familiarity with the given topic of a reading set does not necessarily imply that the examinee cannot score high on the reading set (ETS, 2007). This is because, according to ETS (2007), the way reading sets are developed, formatted, and selected for each single administration of the test provides each test taker with an equal chance to perform well on the reading section as a whole.

Effective versus Ineffective Strategy Use

Variation in strategy use, including efficacy of strategy use, can explain variability of test performance as evidenced in test scores (Bachman, 1990; N. Anderson, Bachman, Perkins, & Cohen, 1991; Phakiti, 2003). Strategy use can be said to mediate between test taker characteristics and test performance, and so can optimize the extent to which test performance hinges on test taker characteristics (Yien, 2001). In their study of test-taking strategy use on the TOEFL-iBT reading tasks, Cohen and Upton (2006) suggested the possibility of judging aspects of effective versus ineffective strategy use among test takers by examining differences between strategies that contributed to the selection of the right answers and those which did not. Researchers (e.g., N. Anderson, 1991; Nikolov, 2006) concluded that effective strategy use is determined by strategy

compatibility with task or item format and suitability of strategy sequencing and application. Experts in learner strategies (e.g., Neil Anderson, Anna Chamot, Martha Nyikos, Rebecca Oxford, Joan Rubin, Larry Vandergrift) are almost all in agreement that the manner in which strategies are clustered or combined better differentiates between effective and ineffective strategy use than does the focus on independent strategies (Cohen, 2005).

Theoretical Framework

Statement of the problem.

As pointed out earlier, previous research has had little to say about how strategy use relates to language tasks specific to a given test of language proficiency. This is especially true for the reading section of the TOEFL-iBT. Exploring strategies specific to language tasks on a specific test is expected to be a worthwhile endeavor en route to forming a more precise picture of how the use of certain strategies as opposed to others can aid test performance on these tasks (Cohen, 2006). Future research representing this move has been instigated to aim at specific populations of examinees in order to draw more valid inferences specific to each population (Phakiti, 2008; Rupp et al., 2006).

As far as the TOEFL-iBT reading section is concerned, there remain research questions to be answered as to how examinees with discrete levels of proficiency perform the reading tasks, what distinguishes strategies each group is more disposed to use from the other, and how the strategies used by each group fit on the scale of effective strategy use for the test. On language tests, proficiency is expected to correlate with test performance, which suggests that proficiency also correlates with efficiency of strategy use which, in turn, correlates with test performance. This implies that our efforts to

discern which strategies high-proficiency or high-scoring test takers tend to use as opposed to low-proficiency or low-scoring testees can have exceedingly practical merits. For example, we can infer that those strategies the former group uses are more effective and so more conducive to test performance than those of the latter (Cohen & Upton, 2006).

In connection with the discussion of strategy use in relation to task formats and proficiency levels, another question can be posed as to what aspects of effective test-taking strategy use can be employed with which tasks on the reading section of the TOEFL iBT. Despite the fact that Cohen and Upton (2006) related strategies used by their respondents to the task and item types on the TOEFL-iBT reading subtest with somewhat less specificity, as mentioned earlier, the researchers did not attempt to explicate aspects of effective strategy use among their respondents. In fact, this would require that strategies be first linked to reading tasks, and then only those strategies that are associated with correct answers to task items, or their combinations, be given more weight in answer to the question of efficacy of strategy use. The answer to this question can gain even more support if these strategies, or their combinations, are shown to be used at higher rates by highly-proficient or high-scoring test respondents when compared to low-proficiency or low-scoring test respondents. It is worth mentioning here that, throughout this research report, reference to proficiency and scoring levels implies levels of reading proficiency and scoring as determined by means of given reading tests. Certainly, scoring levels can be equated with, and so are referred to as, performance levels.

Purpose of the study.

With the focus on how a sample of 25 Arab ESL learners behave in response to the reading tasks on the TOEFL iBT, this research aimed to: first, find out what test-taking strategies respondents use when responding to the task items; second, investigate if there are differences between high- and low-scorers among respondents in their choice and use of test-taking strategies on the task items; and third, determine what aspects of effective test-taking strategy use respondents employ on the task items.

Significance of the study.

There have been scholarly voices calling for more research with the aim of exploring the relationship among strategy use, test format, language proficiency, and test performance (e.g., Cohen, 2006) with participants who have common backgrounds (e.g., Phakiti, 2008; Rupp et al., 2006). Other scholars (e.g., Cohen & Upton, 2006) have encouraged research efforts to inquire into aspects of effective strategy use, including what and how test-taking strategies can be effective in relation to the task types and items on the TOEFL-iBT reading section. On the basis of Cohen and Upton's (2006) study of test-taking strategies on the TOEFL-iBT reading section, among respondents whom the researchers considered highly proficient, it can be assumed that since these respondents used more test-management than test-wiseness strategies, respondents who are at a lower level would use more test-wiseness than test-management strategies.

In response to the above-mentioned research motives and assumptions, this study sought to focus on test-taking strategy use on the reading section of the TOEFL iBT among Arab ESL learners. Given that the population of this study shared linguistic and cultural backgrounds to a great extent, the effect of those factors that were not directly

linked with the test-taking experience on strategy use and test performance was kept to a minimum. Through a study of strategy use among Arab ESL learners when they take the reading subtest of the TOEFL iBT, it would be possible to know what strategies are typically used by respondents with the specific task and item types. Secondly, determining potential differences between high- and low-scoring test takers among respondents would enable us to learn what response behaviors and strategies characterize high test performance, and which act to the contrary. And thirdly, relating strategy use by respondents to their right answers to item types would help us ascertain aspects of effective strategy use on the item types.

It is hoped that the study of Arab ESL learners' choice and use of test-taking strategies in response to the TOEFL-iBT reading section will result in insights about how this population of examinees approaches and performs the reading tasks on the test. Such insights would enable us to propose practical implications for classroom practice, prospective test takers, and test preparation programs. The import of this exploration of test-taking strategies becomes more evident when we consider the amount of effort test preparation and coaching schools exert in order to train examinees in using test-taking strategies as well as the cost of attending these programs for examinees themselves. By and large, such strategies have gained support from the intuitions or unsystematic observations of test takers, rather than insights drawn from empirical research. On the other hand, test takers who have not had training in using test-taking strategies may assume that they have to take the test in the manner dictated by the question formats without attempting to strategize the response process.

Research Questions

With the intention to address the aforementioned research goals with a sample of 25 Arab ESL learners, this study sought to answer the following questions:

1. What test-taking strategies do subjects use when responding to the reading tasks and items of the TOEFL iBT?
2. Are there any differences between high- and low-scorers among subjects in their use of test-taking strategies on the reading tasks and items?
3. What aspects of effective test-taking strategy use do subjects tend to employ with the reading tasks and items?

Assumptions

The study relied upon the following assumptions: (a) the TOEFL-iBT reading sets used in this study called upon an assortment of test-taking strategies the participants would typically use on the actual reading section of the TOEFL-iBT; (b) the participants responded to the reading sets to the best of their language ability and test-taking skill; and (c) the participants reported their test-taking strategies and overall response behaviors accurately and truthfully.

The Researcher's Role

The researcher adopted the emic perspective which is typical of qualitative studies. That is, participants' own views and experiences were taken into accounts in all phases of research, including data analysis and interpretation. The researcher set as his main goal to use unstructured interviews with participants in order to collect as much rich data as possible. This was achieved through the use of questions and prompts that arose during the interview, and were not predetermined by the researcher. Thus, data to be collected using this approach would not be contaminated by any *parti pris* or prior

perceptions on the researcher's part (Mackey & Gass, 2005). A potential caveat of this procedure may be halo effect which refers to an inclination of the interviewee to interact with the interviewer using his intuitions about what the interviewer might be interested in. However, this caveat was overcome through the use of stimuli that helped direct participant's focus to his actual strategy use. (More details on the researcher's role in this study are provided in Chapter 3, under the section of Data Collection Procedures).

Delimitations

The delimitations or boundaries imposed by the researcher on the scope of this study were assumed to control the effects of some external factors upon validity and reliability of research findings and conclusions. First, the study had as its target population ESL learners who speak Arabic as their mother tongue. This served three purposes: to focus the research on a more specific population, to control for sizable effects of L1 on L2 reading had participants from different linguistic backgrounds been recruited, and to encourage participants to express their strategy use without being confounded by any difficulties or problems associated with their use of L2. Since the researcher speaks Arabic, sharing a language with participants helped in three ways: first, any chances of misunderstanding or misinterpretation were minimized; second, participants were familiarized quite easily with the technical features of the test (e.g., reading set instructions, window display, navigation through question items, etc.); and third, all information necessary to meeting the research goals was elicited in as accurate and detailed a manner as possible.

Second, participants were recruited only if they had had the experience of either taking the TOEFL iBT or preparing to take it. The researcher thought that this

delimitation was important to ensure that participants were familiar with the task and item types that appear on the reading section of the test. Therefore, there was no concern that participants' levels of test performance might have been influenced by any factor other than participants' level of proficiency and/or skillfulness of dealing with the reading tasks.

Limitations

There are a few limitations that may affect the representativeness and generalizability of the findings and conclusions reached in this study. Data of this study were collected from 25 ESL learners who came from the same language background (i.e., Arabic), and so may not be representative of the whole population of the TOEFL-iBT test takers. There are two factors that affect generalizability of findings and conclusions of this study to the entire population of Arab ESL learners who have taken the TOEFL iBT. First, this study had 25 Arab ESL learners as participants from a single academic institution; and second, the study employed a convenience sampling. Both factors entail that the findings and conclusions of this study may not be generalizable, without reservation.

Organization of the Study

This study is organized into five chapters:

- Chapter I presents the background of the study, definition of key terms, theoretical framework (including statement of the problem, purpose of the study, and significance of the study), research questions, assumptions, the researcher's role, delimitations, and the limitations of the study.

- Chapter II offers a review of the literature, which includes reading comprehension in the TOEFL context, the TOEFL and assessment of reading comprehension, design and format aspects of the TOEFL reading section, research on validity and performance of the TOEFL reading section, and test-taking strategies on L2 reading comprehension tests.
- Chapter III describes the methodology of this research, which includes rationale for the research design, recruitment of participants, instrumentation, data collection procedures, pilot study, data transcription and coding, and procedures for data analysis.
- Chapter IV presents the results and findings of this study according to research questions, in terms of test-taking strategy use across task and item types, test-taking strategy use among high- and low-scoring groups of test takers across task items, and aspects of effective test-taking strategy use across task and item types.
- Chapter V provides a summary of the study, conclusions, implications for practice, suggestions for further research, and concluding remarks.

CHAPTER II

REVIEW OF LITERATURE

This chapter presents a review of the literature pertinent to the discussion of test-taking strategy use on the TOEFL reading section. The scope of this review has encompassed all key variables expressed in the research goals and questions of this study. More specifically, the discussion here deals with reading comprehension as it is viewed in the TOEFL context, the approach followed in the TOEFL to assessing reading comprehension, the design and format aspects of the TOEFL reading section, and strategy use on reading comprehension tests, including research that has targeted the TOEFL reading section. (The reader is reminded that this review uses the acronym TOEFL without an ending to refer to the test in general or its framework. However, wherever there is a need to refer to a specific version of the test (e.g., TOEFL-iBT), an ending is attached to the acronym TOEFL to reflect the given version).

In light of the research goals and questions stated in the Introduction, this chapter aims to acquaint the reader with the nature of test-taking strategies that test takers tend to use in relation to various formats of reading tasks, aspects of differential use of test-taking strategies by high- and -low proficiency readers on a variety of reading tasks, and aspects of effective versus ineffective use of test-taking strategies and how this relates

to test performance. The chapter is organized into five sections: (a) reading comprehension in the TOEFL context, (b) the TOEFL and assessment of reading comprehension, (c) the design and format aspects of the TOEFL reading section, (d) research on validity and performance of the TOEFL reading section, and (e) test-taking strategies on L2 reading comprehension tests.

Reading Comprehension in the TOEFL Context

The discussion of reading comprehension in the context of the TOEFL is made here in consideration of underlying skills and types of reading, theories and perspectives that sought to explicate reading as a skill and a process, and the interactive view of reading comprehension.

Underlying skills, perspectives, and types of reading.

To begin with, reading is generally seen by specialists as a composite of abilities or skills (e.g., Davies and Widdowson, 1974; Grabe, 1986; Munby, 1978) that can further be classified into *lower level* and *higher level* (e.g., Davies and Widdowson, 1974). *Lower-level* skills involve the ability to comprehend explicit information, while *higher-level* skills entail the ability to figure out implicit information (Grabe, 1986; Schedl, A. Gordon, Carey, & Tang, 1996). Whether or not these sub-skills are separate from one another has been a contentious issue, even with findings drawn from experimental research (Gomez, Noah, Schedl, Wright, & Yolkut, 2007).

For example, Lunzer, Waite, and Dolan (1979) found that the existence of higher and lower levels of comprehension was not supported by their data. The researchers did not observe any significant differences in the way their subjects carried out higher level reading comprehension tasks versus lower level ones. Consequently, Lunzer et al. (1979)

concluded that reading comprehension is not made up of constituent skills, but is a unified construct. In contrast, Alderson and Lukmani (1989) replicated Lunzer and his colleagues' (1979) findings with ESL students whom were instructed to respond to a reading test. Then, the reading skills the students exhibited in their responses were presented to reading specialists to evaluate the use of these skills. The evaluators were able to classify 14 out of a total of 41 question items as assessing lower, middle, and higher level skills. They also found that the lower-level and higher-level question items were shown to assess different skills in that the former tested linguistic abilities, while the latter assessed "cognitive skills, logic, reasoning ability, and so on" (p. 268).

According to Enright, Grabe, Koda, Mosenthal, Mulcahy-Ernt, and Schedl (2000), three perspectives have attempted to offer conceptualization of reading comprehension, which are the *processing perspective*, the *task perspective,* and the *reader perspective*. The *processing perspective* focuses on the differences that exist between skilled and unskilled readers in terms of factors pertinent to textual processing such as automaticity in word recognition, depth of vocabulary knowledge, efficiency of working memory, facility of syntactic parsing, and integration of information (Carpenter, Mikaye, & Just, 1994; Perfetti, 1997). The *task perspective* assumes that reading mainly constitutes tasks performed by readers in order to achieve a particular purpose or goal of their reading and that reading is influenced by task variables that can bring about individual differences in reading performance (Enright et al., 2000). However, one problem that Enright et al. (2000) identified about the *task perspective* lies in its obvious lack of adequate account of textual processing factors and how they relate to task-based reading performance.

The *reader-purpose perspective* addresses how the fulfillment of the purpose of reading can be facilitated through a combination of what the reader knows and how the mind goes about processing what is being read (Enright & Schedl, 2000). The *reader-purpose perspective* view has been seen by experts like Enright et al. (2000) as the best among the three perspectives to implement in reading test design and the one most attuned with the two other views of reading comprehension. It is for ease of interpretability and compatibility with the processing- and task- based views that the *reader-purpose perspective* has been adopted by Enright et al. (2000) as the working rule or principle in the TOEFL 2000 reading framework.

According to Enright et al. (2000), there are four types of reading that ought to be taken into account by any reading theory: *rauding*, *reading to learn*, *search reading* and *reading of multiple texts*. *Rauding*, a term originated by Carver in 1997, *represents* the kind of reading that aims to identify the key concepts and ideas in a text (i.e., basic comprehension); *reading to learn* distinguishes itself from *rauding* by focusing on more specific details; *search reading* stands for the quick reading of a text or table in order to find the key points of information; and *reading of multiple texts* aims to integrate the content of more than one text (Carver, 1997; Guthrie, 1988; Perfetti, 1997).

L1 and L2 reading processes.

The title of this section should not be regarded as suggestive of the fact that L1 reading contrasts with L2 reading; rather, while it is true that there are differences between the two processes, there are similarities that can cause the two to have a reciprocal relationship. The differences between L1 and L2 reading processes can be summarized as follows: L2 reading is influenced by L1 reading background, L2 reading

can be seen as a “cross-linguistic process”, and L2 readers may learn to read in L2 before they have developed adequate L2 oral proficiency (Enright et al., 2000, p. 7). As such, these differences can be said to bring about both facilitative and debilitative conditions for L2 readers; nevertheless, whether L1 reading facilitates or debilitates L2 reading largely depends on the degree to which their two respective languages share typological characteristics pertinent to reading (Enright et al., 2000).

Enright et al. (2000) made mention of four phenomena that can be anticipated with respect to how L2 reading relates to L1 reading:

- *L1 reading aspects are transferrable to L2 reading*. Such aspects include metalinguistic and morphosyntactic elements (e.g., Gundel & Tarone, 1983; Rutherford, 1983), reading skills (e.g., Devine, 1987, 1988), and reading strategies (e.g., Cohen, Olshtain, & Rosenstein, 1986).
- *L1 facilitates L2 reading when L1 and L2 are structurally similar*. For example, facility of lexical processing depends to a large extent on how the structures of L1 and L2 are analogous (e.g., Muljani, Koda, & Moates, 1998).
- *L1 and L2 interact during L2 reading*. For example, differences in L2 readers’ awareness of L2 lexis can be attributed to differences in their L1 lexical processing (e.g., Koda, Takahashi, & Fender, 1998).
- *L2 reading is constrained by L2 low proficiency*. Intra-sentential processing is significantly determined by the level of morphosyntactic awareness which characterizes highly proficient L2 learners (e.g., MacWhinney & Bates, 1989). Also, highly proficient L2 learners learners show faster processing of L2

linguistic input than their counterparts with low proficiency (e.g., Haynes & Carr, 1990).

Based on other findings (e.g., Clarke, 1979; Horiba, 1993), Enright et al. (2000) maintained that L2 reading comprehension can noticeably be restricted in that L2 readers who demonstrate a low level of processing skills cannot grasp as much information from the text and simultaneously integrate information as can L1 readers. It follows from this that L2 reading cannot be equated with L1 reading not only because L1 reading experience continually exercises an influence on L2 reading, but also L2 reading itself is affected by the level of L2 proficiency in its broad sense.

Reading comprehension as an interactive process.

As is the case with other language skills, reading has benefited from the advancements in the theory and practice of language learning that have taken place in company with the communicative competence perspective. In keeping with the communicative view of competence, reading has become more contextualized, taught and learned in relation to both social and educational aspects of its given context (J. Green, 1987). According to Hudson (1996), reading is basically an interactive process, involving the readers' purpose and the context variable. Therefore, readers make use of various strategies that aid their comprehension in accordance with the reading rules imposed by the given context in order for the goal of reading to be more attainable (Hudson, 1996).

As a result, there has been a shift in the earlier views of reading as an activity involving a variety of processes that can be either bottom-up (e.g., character identification, sound representation, lexis processing) or top-down (e.g., use of schemata, text interpretation, inferencing) to the view that reading is an interaction of both

categories of processes (bottom-up and top-down) (Hudson, 1996). There is yet a dichotomy as regards whether the interactive view of language should be ascribed to the reader-text relationship or the way reading processes and abilities relate to one another (Grabe, 1991). However, the reading model that TOEFL 2000 has adopted has taken the interactive view to apply to both (Hudson, 1996).

To sum up, the concept of reading comprehension has been approached from three different perspectives: the *processing perspective*, the *task perspective*, and the *reader-purpose perspective*. However, the latter view has proven to be the most explanatory of reader performance and the most compatible with the other two views; consequently, it stands as the most valid view to adopt in the test design of the TOEFL reading section. According to the *reader-purpose* perspective, reading can best be described as an interaction between the reader and the text, aiming to achieve a particular goal or purpose in the readers' mind and/or the author's message. This interaction has to address the contextual factors without account of which the reader's accomplishment of the reading goal or purpose is restricted, if not hindered. The next section shows how the TOEFL's view of reading comprehension has led to assessment of this skill in a manner specific to the TOEFL context.

The TOEFL and Assessment of Reading Comprehension

The TOEFL is discussed here with respect to the various stages it has gone through until it has reached its current state and form, as well as how it has framed the assessment of reading comprehension.

Inception, evolvement, and evaluation of the TOEFL.

In terms of format, the TOEFL has evolved through three stages: first, it was administered in the form of a paper-based test; next, it was transformed into a computer-based test; and currently, the TOEFL is administered worldwide via the internet. The last version of the TOEFL, in use since September 2005, would not have been a reality without the tremendous amount of research and effort on the part of the ETS and other involved parties. The TOEFL Committee of Examiners, composed of the TOEFL Board and experts from the fields of language education and testing, has served to inform this development process since the earlier 1990s.

Several constituencies with a shared interest in the TOEFL have called for designing an improved version of the test that "(1) is more reflective of communicative competence models; (2) includes more constructed-response tasks and direct measures of writing and speaking; (3) includes tasks that integrate the language modalities being tested; and (4) provides more information than current TOEFL scores do about international students' ability to use English in an academic environment" (Jamieson, Jones, Kirsch, Mosenthal, & C. Taylor, 2000, p. 3). The outcome of these early efforts was the fine-tuning of the TOEFL-CBT, which has later brought about the development of the TOEFL iBT. The TOEFL in its new format, or the TOEFL iBT, can now be delivered to test takers wherever the internet is accessible. With the addition of speaking to the iBT version, the TOEFL now has four sections including listening, speaking, reading , and writing, each of which is assigned a score between 0 and 30, with the total score of the test being 120 (Zhang, 2008).

The validation process of the TOEFL has been ongoing since the inception of the test, and more noticeably, since the initial attempts were made to design the 2000 version (Chapelle, Enright, & Jamieson, 2008). Since validity speaks to how the interpretations we lend to test scores are supported in light of the test uses (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), the interpretations of the TOEFL-iBT scores were sought to reflect the test taker's ability to use academic English. This is expected to aid decisions about a given test taker's admission to an academic program of study or placement in an intensive language program (ETS, 2011a). It was especially important not to directly deal with the individual test tasks in isolation from the abilities they aim to represent, since one task type may test several abilities (Gomez et al., 2007). (See Appendix A for a layout of the TOEFL 2000 validity argument in terms of propositions about how the test scores are interpreted and used, as well as evidence supporting each one of these propositions).

ETS (2011a) maintained that there is evidence suggesting that the reliability and comparability of the test scores are constant across its administrations. The reliability estimates for the test sections are within acceptable limits, ranging from 0.74 to 0.88. Validity exhibits itself through several facets each of which needs to be a characteristic of the test. Thus, at every single step of the TOEFL validation process, caution has been taken to ensure that the test shows evidence of the various facets of validity. For example, content validity has been established through analyzing and identifying academic tasks typical of those performed at the university level and then figuring out language aspects that are used to carry out these academic tasks (C. Taylor & Angelis, 2008).

Recent studies (e.g., Biber and his colleagues, 2004; Cumming, Grant, Mulcahy-Ernt, & Powers, 2005; Rosenfeld, Leung, & Oltman, 2001) have lent strong support to content validity and authenticity of the TOEFL. The work of Biber et al. (2004) was more focused on the listening tasks and whether they are representative and authentic. Cumming et al. (2005) were interested in the integrated tasks that include speaking and writing. In the Rosenfeld et al. (2001) study, undergraduate and graduate faculty and students in a variety of institutions of higher education in the United States were surveyed as to whether the tasks used on the TOEFL are important to academic study at the university level. The findings from these studies support the use of the tasks that are now part of the TOEFL iBT.

Task design and scoring rubrics have been attended to through several research studies to ensure high quality not only of the test design, but also of the mode of test administration and scoring. Such studies have been informative in the decision-making process with regard to the number, types, characteristics, and time allotment of the testing tasks, the interface used to administer the test, and other instrumental and technical aspects of the test administration (ETS, 2008). Furthermore, experts' recommendations (e.g., Brown, Iwashita, & McNamara, 2005; Cumming, Kantor, Baba, Eouanzoui, Erdosy, & James, 2006) have been taken into consideration in the development of scoring criteria for the speaking and writing sections (ETS, 2011b). Other studies (e.g., Brown et al., 2005; Cohen & Upton, 2006; Cumming et al., 2006) with focus on test-taking strategies and processes substantiated the expectations of the test maker regarding the validity and quality of the testing tasks (ETS, 2011b). The results of these studies have

shown that while responding to the typical TOEFL tasks, test takers engage in the same strategy use and processing associated with the response to these tasks in academia.

Perhaps not less important, these studies, specifically Cohen and Upton (2006), ruled out the possibility that test takers could rely on test-wiseness in their efforts to obtain high scores on the TOEFL; that is, test takers were found to use more test-management strategies than those of test-wiseness (ETS, 2011b). As for the test structure, factor analytic studies (e.g., Sawaki, Stricker, & Oranje, 2008) have shown that the TOEFL is composed of one main factor accounting for the English language ability, in addition to four factors related to the four skills that make up the test. ETS (2011b) noted that this finding attests to the high consistency of the test with respect to the representation of the ability underlying constructs, as has empirically been confirmed with a variety of language groups (e.g., Stricker & Rock, 2008).

As noted earlier, test scores are essential to the validation process of a test. Also, whether or not a given test is reliable can mainly be determined by how consistent its scores are across multiple administrations of the test. The TOEFL scores have been shown to correlate with the results of other measures of language proficiency including self-assessment (Wang, Eignor, & Enright, 2008), academic placement (Wang et al., 2008), institutional local tests for international teaching assistants (Xi, 2008), and performance on simulated academic listening tasks (Sawaki & Nissan, 2009). ETS makes every effort to capitalize on the positive consequences and to minimize the negative ones associated with the interpretations and uses of the TOEFL scores (ETS, 2011a). In keeping with what Messick (1989) suggested regarding the advantage of assigning interpretations to test scores so as to make them of expressive values to testees and test-

score users, those parties who are interested in the TOEFL scores are provided with explanatory and interpretive information about these scores (ETS, 2004, 2005). Not only does ETS lend priority to admission and placement decisions linked to the performance on the TOEFL, ETS is also hopeful that this test aids in the proliferation of the quality of the programs that offer training in the English language and/or preparation for the TOEFL.

Frameworks for assessing reading comprehension in the TOEFL.

The theory and research on the TOEFL over the last two decades have informed the test framework and design, and have brought about the revolutionary iBT version of the test. In the first attempts to lay out the test design, it was postulated that the *reader-purpose perspective* is the best model to follow for a number of reasons: First, test-score users are readily offered interpretation of scores that can guide their decision-making. Second, the *reader-purpose perspective* shows high compatibility with the other two perspectives, namely the processing- and the task-based views of reading (Enright et al., 2000). In their proposed framework for the TOEFL 2000 reading section, Enright et al. (2000) recommended that reading comprehension be assessed through the engagement of test takers in four purposes pertinent to academic reading, which inherently relate to the processing- and task-based perspectives. These four purposes can be summarized as follows.

1. *Reading to find information:* readers try to identify explicit information in order to acquire knowledge, answer questions, or clear delusions. Proficient reading depends on how automatic and efficient bottom-up reading skills are. Reading tasks require readers to look for specific information.

2. *Reading for basic comprehension:* readers try to figure out the implicit main idea or major points of the reading. Proficient reading at this level depends on how readers are able to form a mental version of reading promptly and competently. Reading tasks call for readers to deduce the main idea or topic, or distinguish between major and minor points.
3. *Reading to learn:* readers focus on how to relate discrete pieces of information to one another to have integral understanding of the reading. Skilled reading is determined by the reader's ability to figure out the nature of the relationship that links different pieces of information together such as cause-and-effect, comparison and contrast, sorting or classification, or argumentation or persuasion. The reading tasks at this level call for constructing a theoretical representation of the reading which offers an interpretation of its content.
4. *Reading to integrate information:* readers direct their attention to the integration of the content of two or more textual formats. Proficient reading is largely dependent on the reader's ability to form integral comprehension of the information presented in these texts. The reading tasks for integrating information require the reader to offer a unified, conceptual framework of the content of two or more texts (pp. 5-6).

Enright et al. (2000) suggested that these four purposes of academic reading touch upon the construct of reading comprehension from different angles by each calling for the use of such abilities as word processing and comprehension with some variation between the first pair (i.e., *reading to find information* and *reading for basic comprehension*) and the second one (i.e., *reading to learn* and *reading to integrate information*). The authors

added that we could place the four purposes along a difficulty continuum ranging from *reading to find information* to *reading to integrate information,* with the latter being the most complex and demanding of all four purposes.

Taking into consideration the nature of the L1-L2 typological relationship as being either convergent or divergent, Enright et al. (2000) devised two important implications for L2 reading assessment. First, since L1 potentially facilitates or debilitates L2 reading comprehension performance, L2 readers from different language backgrounds are more likely to exhibit varied levels of performance on L2 reading assessment tasks. This necessitates that care be taken with the time allotment of the test and the representation of L2 reading skills and abilities in the test content. And second, research on early L1 and L2 reading development has shown that L1 reading benefits from the reader-text interactive view, whereas L2 reading relies upon bottom-up processing. This suggests that L2 reading assessment tasks ought to be constructed in accordance with the processing view shown by research to be appropriate to a given level of L2 readers.

As mentioned above, the interactive view of reading has been adopted in the TOEFL reading section for the reason that this view has stemmed from the communicative model of competence. According to Hudson (1996, pp. 4-9), the current view of reading as an interactive process has been tailored to consider the following aspects in both L1 and L2 reading, which are of marked import to the performance and assessment of academic reading:

- *Automaticity in word and sentence recognition.* Reading comprehension is essentially reliant upon how fast and effective its underlying processes and

abilities (Pollatsek & Rayner, 1990). For example, a strong correlation has been shown to exist between how efficient such abilities as lexical processing and decoding abilities are and reading comprehension in L1 (Cunningham, Stanovich, & Wilson, 1990). Furthermore, the extent to which such abilities can be efficiently performed depends on the extent to which the reading activity is contextualized (Balota, 1990). Word recognition has proven to be an important bottom-up process, as shown in studies which reported that proficient readers tended to catch sight of a great deal of textual information (e.g., Drum, Calfee, & Cook, 1981; Just & Carpenter, 1987; Pollatsek & Rayner, 1990). Skilled processing of grammatical relations has also been shown to correlate strongly with the amount of textual information processed while reading (Drum, Calfee, & Cook, 1981; Haynes & Carr, 1990).

- *Content and formal schema*: (a) *Content schema.* The extent to which the subject matter of the reading is familiar determines its comprehensibility. When readers read, they try to comprehend by relating the new pieces of knowledge to their background knowledge (R. Anderson & Pearson, 1988). Readers tend to use their background knowledge as a guide to locate particular information they need (Symons & Pressley, 1993). Bower, Black, and Turner (1979) found that when readers were asked to retell a story in which they had read about a familiar experience, they were inclined to elaborate by referring to their prior knowledge. (b) *Formal schema.* The reader's level of familiarity with textual and organizational features of a certain text determines the level of text comprehensibility. Reading rate and comprehension were found to be more

associated with texts that make enough use of devices denoting anaphoric and cataphoric associations (Irwin, 1986). This is because such devices contribute to cohesion which, in turn, promotes coherence (Van Dijk, 1977). Readers tend to excel when reading texts that show more coherence than those that do not, even when both texts present the same content (McKeown, Beck, Sinatra, & Losterman, 1992). On the other hand, texts with features that conflict with coherence were found difficult to comprehend (Beck & McKeown, 1989). The way texts are structured and organized in a logical and hierarchal manner can direct and facilitate both comprehension and recall (Carrell, 1984, 1985), more noticeably, when readers are acquainted with the text (Roller, 1990).

- *Strategies and metacognitive skills.* As readers read, they make use of a wide range of strategies related to their reading ability, which helps them get the most out of what they read. By looking into strategy use, we can gain insights as to how readers interact with the materials they read and how they manage to comprehend (Carrell, 1989). Barnett (1988) grouped reading strategies into word-level strategies and text-level strategies. The former are used when dealing with words (e.g., guessing the meanings of unfamiliar vocabulary), while the latter are used when dealing with a portion of the text or the whole text (e.g., predicting what the text is all about from reading a few of the first sentences). Block (1986) counted as qualities of skilled readers: the ability to track the flow of understanding, the decision-making as to what strategies are needed, and the knowledge of how to use those strategies.

- *Purpose and context.* In view of communicative competence and the *reader-purpose perspective*, writers of academic texts should see reading in this context as interactive, and so strive to make it more of a meaning- and purpose-based process. An academic text ought to represent the medium of social interaction between its author and its reader despite the fact that for each the text would serve a certain purpose (Hill & Parry, 1992). The purpose and the context of reading can bring about varying levels of reading performance even when readers are at the same level of reading ability.

Messick's (1989) unitary view of validity has been taken care of in the TOEFL test design since the meeting of the TOEFL Committee of Examiners in the spring of 1993 (Hudson, 1996). In his view, Messick (1989) maintained that validity can best be seen as an integral notion that represents, first and foremost, construct validity along with its six underlying facets that include: 1) relevant, representative, and appropriate content; 2) theoretical and substantive evidence of construct representation; 3) construct-relevant and -consistent scoring rubrics; 4) generalizable score interpretations and inferences; 5) evidence of criterion-related validity that derives support from measures of convergent and discriminant validity; and 6) an appraisal of social and educational consequences of test uses and scores. The essentiality of construct validity in the TOEFL 2000 reading framework has been claimed by Enright et al. (2000), who noted that construct representation is used by Messick (1995) to refer to two notions: First, *representativeness* which describes the degree to which the assessment represents the content being assessed as indicated by its scope of coverage. And second, *representation* which reflects the extent to which a given assessment task requires a psychological involvement of each of

the ability being assessed, the task distinguishing features, and the cognitive processes and strategies that the test taker uses when carrying out the given task.

On the basis of the above-highlighted aspects that demand attention in L1 and L2 models of reading, Hudson (1996) proposed a set of implications for assessing academic reading that align with the Messick unitary view of validity. He cited other scholars to support his implications. Such implications can be summarized as follows:

- In order to ensure adequate representation of the academic reading construct, tasks eliciting constructed responses ought to be included on reading tests in addition to selected-response ones. This is because selected-response tasks do not actually reflect real-world reading where readers have other means to demonstrate their understanding and knowledge of what they read (Mislevy, 1993). Also, selected-response tasks relate to the misassumption that reading skills are more fragmentary than they are in reality, especially when we want to deal with reading as a unitary process involving its underlying skills in a synchronized, overlapping relationship (Bennett, 1993). Besides, positive washback can be brought about when moving from assessment to instruction by teachers' assigning a variety of reading tasks that call for constructed responses instead of those designed to prepare learners to deal with selected-response formats.
- Despite the above-mentioned shortcomings of selected-response tasks, they are still useful to the assessment of general and non-context specific skills of academic reading. This has been emphasized in order that general skills and context-specific skills are well-balanced in the context of academic reading (Bennett, 1993).

- With the aim of assessing the various skills underlying academic reading, following a task-based approach in the test design helps provide more representation of these skills (Long & Crookes, 1992). Tasks used for this purpose need to be similar to those typical of academic reading to ensure authenticity of such tasks. The amount of textual information presented in a reading needs to be rich enough to allow for ample reflection of the context it intends to depict (Peirce, 1992). This is expected to satisfactorily meet the standards of fairness and equity linked to the use of the given test. In other words, when the test is authentic and reflective of reading academic tasks, it becomes less likely that this test will advantage a particular group of test takers at the expense of others.
- The texts used on reading tests should be selected such that they are based on those themes typically referred to and dealt with in the academic world. These texts can be said to be authentic in as much as their surface topics and supporting content possess features of real-life academic issues. When there are a variety of texts chosen on the basis of academic themes, test takers can be allowed the choice among these texts as per their academic fields of study. This can remarkably be facilitated through the use of computerized tests.
- As part of the fulfillment of consequential validity, test-score users should be provided with test scores accompanied by guides or manuals that can help interpret the scores, such as the use of descriptive scales. This can also help answer certain questions that the end test-score users may have as regards the strengths and weaknesses of candidates applying for a job or to a program of study. Such descriptive information about scores can also guide and assist with

decisions about placing language learners at levels matching their abilities, or diagnosing any kind of weaknesses or problems they may have.

- Academic reading can be integrated with writing through the use of structured-response tasks. For example, test takers may first read a passage and then engage in writing a summary of what they have read. Clearly, reading and writing can be combined in a single test administration; and so, overall academic literacy can be assessed. (pp. 9-12)

Two points pertinent to Hudson's (1996) implications for the TOEFL 2000 reading framework deserve more stress here before concluding this section. First, providing test takers with descriptive information about their performance is highly encouraged, especially when we consider their struggle for academic development. Test takers hope to get feedback or diagnostic information as to the kinds of weaknesses that have influenced their performance, and the tasks that have posed special problems to them (Gomez et al., 2007). Secondly, another advantage of computer-based testing includes bringing about more creativity and novelty in the test design through the use of technically advanced features, for example, the possibility of incorporating a variety of innovative task formats (Cohen & Upton, 2006).

To conclude this section, it is evident that the current TOEFL represents the outcome of longstanding thought and research on the part of the ETS in cooperation with other agencies and parties. Consultations have been sought with specialists and experts in the fields of both language teaching and learning, and language testing and assessment to ensure the adequacy of any decisions to be made in regard to the test design. In order for the assessment of academic reading by the TOEFL to be proper, it needs to have ample

evidence for established validity in light of the Messick unitary view of validity. It should also incorporate tasks with content that embodies the four purposes of academic reading (i.e., *reading to find information, reading for basic comprehension, reading to learn,* and *reading to integrate information*) in communicatively-oriented situations. It is shown in the following section how the transition has been made from theorizing and researching assessment of reading comprehension to the design of the reading section of the TOEFL.

Design and Format Aspects of the TOEFL Reading Section

Part of the purpose of the previous sections is to lay out the stage for the discussion of certain aspects of the reading subtest of the TOEFL. Such aspects as those of the design and format of the TOEFL reading subtest constitute the focal point of this section.

The goal of the reading section of the TOEFL is "to simulate the types of reading tasks that students are expected to do in university-level academic settings" (Cohen & Upton, 2006, pp. 5-6). As mentioned earlier, *the reader-purpose perspective* has been adopted in the TOEFL 2000 reading framework which, according to Enright et al. (2000), has made academic reading well-defined in the test design by emphasizing more purposes of reading comprehension than those in focus a decade ago. The framework proposed by Enright et al. (2000) for the TOEFL 2000 was implemented by introducing two out of the four purposes of academic reading, namely, *reading for basic comprehension* and *reading to learn* into the design of the reading section (Enright & Schedl, 2000, p. 5).

The design stage.

The design stage for the TOEFL 2000 reading section consisted of four steps (Enright et al., 2000). First, *construct identification* in which the construct of academic

reading was described through a careful review of the literature and operationalized through the identification of tasks and techniques that embody it. Second, *prototyping* which involved the construction of the test tasks along with their respective scoring rubrics that can be modified or improved in light of relevant, concurrent research and psychological measurements. Third, *pilot testing* in which the prototyped tasks were tried out with small groups of test takers to pinpoint any areas in need of improvement and to discern the degree of the design suitability for future examinees. And fourth, *field testing* in which the tasks that stood the *pilot testing* stage were field-tested with large groups of test takers in order to use rigorous statistical methods to check the overall validity of the test design. This step also involved focus on the individual items and the scoring systems.

According to Enright et al. (2000), the task characteristics of the TOEFL reading section have been identified in the framework in terms of *situation, text material,* and *test rubric*; each has a set of variables operationalized to best reflect the context of the test. For *situation*, participants in the context of academic reading include the reader whose role is to make use of his available cognitive and linguistic resources to comprehend the message and the author whose role can be made clear in accordance with the purpose and the type of the message she intends to convey. While academic reading can be carried out in a variety of settings with different physical and sensorial features, the test design does not have to describe the particulars of these settings unless it is feared that readers' performance may be influenced by certain situational aspects, such as the case of integrative modes of testing.

When selecting the test content, adequate sampling of readings from a variety of disciplines is encouraged provided that test takers are not disadvantaged, for example, by

having them deal with readings from highly specialized areas within these disciplines. The communicative purpose of the selected texts needs to match what test takers expect when they read these texts, as their understanding can be maximized when they know how to relate the purpose of their reading to their background knowledge and strategy use (Goldman, 1997). Register, referring to the use that the author intends the text material to serve, is likely to vary across an array of texts. Therefore, certain parameters of register need to be comparable across test administrations so as to provide test takers with an equal opportunity for optimum performance on the test.

The *text material* on the TOEFL reading test ought to include a selection of texts with varying degrees of length and use of supporting visual features. The inclusion of various grammatical and discourse features, on the one hand, and pragmatic and rhetorical features on the other, needs to be based on certain criteria that both gain support from previous research findings and benefit from the current ones, in an ongoing process of refinement and substantiation. And last, *test rubrics* can take the form of several arrangements of the text (A) and task (B) elements (as seen in the table in Appendix B), which is proposed to be an aspect of the TOEFL 2000 reading test design. Through the use of this scheme of task and text aspects, the test designers can easily match each one of the four purposes of academic reading with a reading task that fits the given purpose. For example, assuming that the intention is to design a *Reading to Integrate* section for a single administration of the test, the tasks that are created may ask test takers to think of a suitable solution to a given problem discussed in a number of texts.

In their framework, Enright et al. (2000) suggested that using a computer-based mode for assessing reading comprehension necessitates the use of certain features so as to make the experience of using a PC interface as similar as possible to the paper-based mode. Accordingly, they strongly recommended the use of a reader-friendly interface by using several reading passages including supporting features (e.g., pictures, graphs, maps, videos, etc.), visible tools for navigation, lucid resolution and display, full-text viewing, tools for scrolling up and down and magnifying the text, and immediate reaction and processing (pp. 41-42). As for response formats, Enright et al. (2000) suggested that in order to assess the four purposes of academic reading, a variety of response formats in addition to multiple choice items can be made possible through the use of computer technology (e.g., constructed response items, selection of a word or part of the text, and completion of a table or chart) (pp. 37-38).

Task specifications.

The TOEFL reading section in its current form incorporates the first three purposes set forth in the Enright et al. (2000) framework in addition to tasks that call for making inferences, which share some features with the other task types (Cohen & Upton, 2006). What follows are the task specifications for the TOEFL reading section according to ETS (2003):

1. *Basic comprehension tasks* measure both lower and higher level abilities with the former being reflective of word-level decoding and the latter of sentence-level understanding.
2. *Inferencing tasks* assess higher level abilities of relating pieces of information at the sentence level to one another such that the reader can deduce some other

information not presented in the text, but relevant to the author's intent. *Reading to learn* tasks measure higher level abilities through which the reader can demonstrate how the textual information is structured, how to hierarchize different segments of information in relation to the topic being discussed, how to build a conceptual schema of the text, and how to relate the details to the main points to fulfill a certain stylistic purpose, for example, cause and effect.

Consequently, there are three task types in the current TOEFL or the TOEFL-iBT which, along with the number of items for each task type, include: Basic Comprehension (five items), Inferencing (three items), and Reading to Learn (two items). The Basic Comprehension five item types are *vocabulary, pronoun reference, sentence simplification, factual information,* and *negative fact*. The Inferencing three item types are *basic inference, rhetorical purpose,* and *insert text*. The new item types comprising the Reading to Learn task are *prose summary* in which test takers are provided with a list of six sentences offering key points from which test takers need to choose three sentences to complete a summary of the text, and *schematic table* which asks test takers to drag a number of sentences, one at a time, and drop them into a table intended to display the major themes of the text (Enright & Schedl, 2000, p. 19). Obviously, the Reading to Learn tasks aim to test the ability to distinguish between major and minor points in order to offer a summary or schematic outline of the reading passage (Cohen and Upton, 2006).

According to ETS (2007), the reading section of the TOEFL-iBT presents examinees with three to five passages similar in content and complexity to those readings used at the university level. Each reading passage has from 600 to 700 words, and its genera can be either expository, argumentative, or historical narrative. The latter can be

either autobiographical or biographical on a single test administration. The TOEFL-iBT reading section has three item formats: a) single-answer four-option multiple choice; b) single-answer four-option insert text; and c) partial-credit reading-to-learn items with six to seven answer choices (Alderson, 2009). (Appendix C provides a brief definition and description of each one of the item types along with an example of how each item type is operationalized in the TOEFL-iBT reading section).

Research on Validity and Performance of the TOEFL Reading Section

Ever since the TOEFL was put into use, a number of studies have looked into a selection of aspects including the test design, tasks, and features. The studies to be reviewed here are those that looked at aspects pertinent to the TOEFL 2000 framework which applies to the three formats of the test (viz., paper-based, computer-based, and internet-based or iBT).

A model of test-taking.

In their efforts to propose a framework for the TOEFL 2000 and to validate both construct representation and score interpretation specific to the reading section, C. Taylor et al. (1998) selected 20 reading question items and a similar number of listening question items from the item bank used for trialing the TOEFL-CBT. Then, they hired a number of test takers and provided them with a reading test composed of a passage and a set of question items on the passage from the 20-item pool. The researchers observed that the examinees mainly followed the Kirsch and Mosenthal (1990) *model of prose and document processing in reading* with a question asking for the selection of one option out of four, which best represents the main topic of the passage.

C. Taylor et al. (1998) noted that the examinees first attempted to use the question to determine the purpose of the question itself and, on this basis, looked for the relevant information in the text that pertained to the question intent. Next, the examinees were careful to attend to the difference between the given information of the question and its requested information, which are certainly not the same. For example, if the given information of the question suggested that there was a main topic of the text, the requested information required that the test takers construct a mental, thematic map of the text to help them figure out the topic. Then, the examinees tried to skim the text and locate the information that matched the given and/or the requested information in the question stem or options. Once they found the matching information in the text, they moved on to the next step; if not, they tried to use inferencing or background knowledge, or returned to a previous step.

After that, the examinees tried to supply the requested information; or else, they had to either return to a previous step or rely on inferencing or background knowledge. Last, the examinees chose to go through the process all over or in part to make sure they had answered correctly; if not, they tried to locate the requested information somewhere else (e.g., by searching the whole document). Jamieson et al. (2000) argued that any reading response format (e.g., open-ended, multiple-choice, etc.) used with a document or prose calls for test takers to follow this model of reading test-taking.

Predictors of task difficulty and test taker performance.

By means of Kirsch and Mosenthal's (1990) *model of prose and document processing in reading,* C. Taylor et al. (1998) found that there are three sources of task difficulty that can aid us in making predictions about the level of difficulty associated with tasks we intend to include on a reading test. These sources or predictors of task difficulty can be described as follows:

- *Type of information requested.* The type of information requested through a question determines how test takers go about answering that question successfully. Questions asking for information pertinent to a concrete quality of an entity are much easier to tackle than those asking for abstract information.
- *Type of match.* The type of match the test taker needs to attend to when attempting to provide the requested information depends on the amount and complexity of both the given and requested information in the question. The content load of the requested information determines how complex the type of match that the test taker is required to make with the information presented in the text in order to provide the correct answer.
- *Plausibility of distractors.* The availability of distractors that seem to share some features with the requested information intrinsically constitutes an additional dimension of complexity that test takers need to overcome. That is to say, the extent to which distractors are plausible is a significant determinant of the complexity level of a given question. (pp. 21-23) In their experimental work, C. Taylor et al. (1998) examined how these three sources of task difficulty determine test takers' performance on reading tasks. In so doing, they performed a

regression test involving each one of these sources (variables) and item difficulty for each one of the 20 question items. They found that 86 percent of the variance in task difficulty was accounted for by the three variables. In another procedure, C. Taylor et al. (1998) used item response theory analyses to figure out how task difficulty relates to test taker ability. They noticed that task characteristics could be used to describe test taker ability and to construct levels of performance when both the test taker ability and the task level of difficulty are close.

Assessment of academic skills.

Another earlier study was carried out by Rosenfeld et al. (2001) who sought to determine the extent to which the TOEFL actually assesses academic skills. The researchers administered two survey instruments of 42 statements about academic skills to 370 graduate and undergraduate faculty and 345 ESL graduate and undergraduate students in 21 American and Canadian universities. The purpose of the survey instruments was for the participants to rate the importance of the academic skills expressed by each task statement for ESL university students in the world of academia and the frequency of use of such skills among the most successful ESL students. The faculty and student respondents rated all 42 task statements as either “important” or “very important” for studying at the university level.

The researchers concluded that the TOEFL makers’ expectations about the functionality of the test tasks are endorsed by the general population of faculty and students at the university level in Canada and the US. As one implication, the researchers recommended a set of reading tasks to be included in the TOEFL. The selection of these tasks was determined by the level of agreement among the respondent groups about the

importance of such tasks for academic success as well as their frequency of use among the most successful ESL learners. Table 1 shows these tasks in relation to the reading skill content domain.

Table 1

Examples of Tasks that Meet Suggested Criteria for Inclusion in the TOEFL Reading Section (Rosenfeld et al., 2001, p. 49)

Content domain	Task statement
Basic comprehension	• Determine the basic theme (main idea) of a passage. • Read and understand written instructions/directions concerning classroom assignments and/or examinations.
Learning	• Read text material with sufficient care and comprehension to remember major ideas and answer written questions later when the text is no longer present. • Read text material with sufficient care and comprehension to remember major ideas.
Integration	• Compare and contrast ideas in a single text and/or across texts. • Synthesize ideas in a single text and/or across texts.

Evaluation of the new task type: Reading to learn.

The Cohen and Upton (2006) study aimed to use evidence from the use of reading and test-taking strategies among 32 test-takers, from four language backgrounds, to judge the usefulness of the *reading to learn* task types. The data on the test takers' strategy use were collected using think-aloud protocols in which each respondent verbalized his or her thoughts while responding to reading tests containing two 600-700 word readings, each with 12-13 question items. The results showed that the *reading to learn* task items did not call for the test takers to use the kind of text processing strategies they were expected to use when dealing with long texts. Rather, the *reading to learn* task items were shown to assess academic reading skills in ways similar to those in which the *basic comprehension*

and *inferencing* task types performed. The results also showed that the *reading to learn* task types were not more difficult than the other two task types as shown by success rate and verbal report data.

The researchers further noted that since the *reading to learn* tasks appear third among the three task types, they did not call for the use of the academic skills they have been designed for. That is, these tasks fell short of assessing the extent to which test takers had a global understanding of the texts. The researchers explained this observation, suggesting that by the time test takers started responding to the *reading to learn* tasks, they had had adequate grasp of what the reading is all about through their dealing with the previous question items in the *basic comprehension* and *inferencing* task types.

Descriptors of reading proficiency.

In another study, Gomez et al. (2007) attempted to develop proficiency descriptors for the TOEFL-iBT reading section by means of scale-anchoring. The researchers' goal was to create descriptions of the abilities necessary for providing correct answers to the questions on the test. Similar to the observation that C. Taylor et al. (1998) made, Gomez et al. (2007) noted that text and task characteristics that form sources or predictors of task difficulty can be used to describe the test taker ability and to construct levels of performance. The scale-anchoring procedure started by establishing a scale which consisted of three equal percentiles, and a set of criteria were proposed to decide about the level at which certain questions can anchor, including measures of item difficulty and discriminability.

Descriptions of the necessary skills and abilities that test takers made use of to answer each question at a given level were obtained from ETS reading specialists. The

analysis, along with the descriptive information about test takers' abilities, were used to identify descriptors of what test takers are capable of at a given level in connection with three question types: *factual information*, *rhetorical purpose*, and *reading to learn*. (For shortened score report descriptors for each one of the three levels along with its respective score range, see the table in Appendix E).

Content knowledge and test performance.

Liu, Schedl, Malloy, Kong, and ETS (2009) attempted to determine if content schemata affect test performance on the reading section of the TOEFL iBT. According to the researchers, this investigation was motivated by the fact that reading passages on the reading section of the TOEFL iBT are fewer than those on the TOEFL-CBT. The researcher maintained that reducing the number of reading passages may give rise to a potential impact of cultural- or educational-related knowledge on performance on the TOEFL-iBT reading subtest. Thus, the researchers examined six passages used in five previous administrations of the test in the years 2007 and 2008. The content of three of the six passages was related to physical sciences, while the other passages revolved about cultural information pertinent to European countries and Japan.

The test takers who took the test in 2007 and 2008 were surveyed about their fields of academic study and cultural affiliations. Two analytical procedures were used: First, differential item functioning was intended to reveal if item function varies with test takers who share the same level of language ability, but belong to different linguistic or cultural groups. Second, differential bundle functioning was aimed to detect if a bundle of items, having one characteristic or more in common, would function differentially with test takers who share the same level of language ability, but represent different linguistic

or cultural backgrounds. The results showed that only few single items and item bundles functioned differentially. The researchers concluded that overall cultural- or educational-related knowledge does not affect performance on the TOEFL-iBT reading subtest.

In summary, the reading section of the TOEFL iBT has gone through four stages of test design in which the construct of reading comprehension was first identified and operationalized in terms of construct-relevant tasks and techniques. Then, scoring rubrics were developed and validated in conjunction with their respective tasks. After that, the test was tried out with small groups of examinees to pinpoint any potential problems in the test design. Lastly, the test was field-tested on a large scale to allow for the use of statistical validation.

As for the test format, the current reading test has three task types which, along with their item types, include Basic Comprehension: *vocabulary, pronoun reference, sentence simplification, factual information,* and *negative fact*; Inferencing: *basic inference, rhetorical purpose,* and *insert text*; and Reading to Learn: *prose summary* and *schematic table.* The last task type represents the innovative vantage of the current format, which is intended to measure test takers' ability in text-level comprehension and to distinguish major from minor ideas.

The research on the TOEFL reading section continues to offer insights about how to enhance the overall test design and format. For example, while *basic comprehension, inferencing,* and *reading to learn* task types have been shown to assess academic reading skills (Cohen & Upton, 2006), it is advisable that each task type call for specific skills, especially the *reading to learn* task type which is intended to assess global understanding of textual information. Thus far, the TOEFL's assessment of reading comprehension has

been explicated in terms of conceptualizing the construct of reading comprehension, identifying useful approaches to the assessment of this skill, and developing adequate test design and task formats for the reading section of the test. The next section presents the theory and research on test-taking strategy use on tests of reading comprehension, with connections being made with the case of the TOEFL reading section.

Test-taking Strategies on L2 Reading Comprehension Tests

The last two decades have witnessed a proliferation of research on both language learning and language use strategies. For the most part, this research has looked into the kinds of strategies that language learners use to facilitate their learning and use of the target language. Not until very recently have scholars and researchers (e.g., Cohen, 2006; Cohen & Upton, 2006; Tian, 2000; Yoshizawa, 2002) become increasingly interested in the kinds of strategies the test takers employ when responding to a form of language assessment. This, in turn, has led to ongoing research to explore the nature of these strategies and how they relate to a multitude of factors involved in the two processes of language test-making and -taking. Hence, this section foregrounds the previous thought and research on the use of test-taking strategies on reading comprehension tests.

Strategies in language learning and language use.

Learner strategies.

In this section, the categories of learner strategies and test-taking strategies are defined and set apart from each other, their defining aspects are described, and those strategies specific to reading comprehension are detailed. It is worth noting here that the term *learner strategies* is used in the literature to refer to both language learning and language use strategies.

Since the 1970s, the interest in the study of learner strategies has started hand in hand with the shift of focus from the outcomes of language learning to the mental operations involved in the learning process (Purpura, 1999). One of the most comprehensive definitions states that learning strategies are "specific actions taken by learners to make learning easier, faster, more enjoyable, more self-directed, more effective, and more transferable to new situations" (Oxford, 1990, p. 8). Extensive research over the last four decades has made it clear that language learners use certain tactics that facilitate their language learning and language use.

According to the majority of scholars, learner strategies represent problem-solving techniques (e.g., Smith, 1979), or conscious tactics used by language learners to compensate for the lack of or deficiency in automatic processes (Phakiti, 2006). There is an overall consensus among L2 strategy scholars that when learners resort to strategies, they deliberate their strategy choice and use and so choose those strategies that they have found through their learning and experience to be facilitative of the task they are performing (Hsiao, 2004). Therefore, strategies are conscious and planned choices that language learners make use of in order to aid their dealings through the target language (Cohen & Upton, 2006; N. Anderson, 1991).

Nonetheless, for some scholars, strategies are actions (i.e., behavioral); for other scholars, strategies are mental processes; and for yet others, strategies can be both behaviors and processes (Ellis, 1994). It seems that this disagreement about whether strategies are observable behaviors or mental processes has to do with whether these strategies are conscious or automatic. However, the view of strategies as behaviors and processes can be qualified on the ground that strategies in the early stages of their use

tend to be conscious and deliberate, and as the learner becomes more experienced and proficient such strategies become subconscious and automatic processes or skills (Afflerbach, Pearson, & Paris, 2008; Phakiti, 2006).

Language learning strategies and language use strategies are two categories of learner strategies with the purpose of the former being to assist the language acquisition process and the latter to aid the performance of language tasks. One of the most influential classifications of learning strategies is the one by Oxford (1990), which includes such strategy categories as *cognitive* (used to control the language), *metacognitive* (used to plan, monitor, and evaluate learning), *memory* (used to remember and retrieve linguistic information), *compensation* (used to repair or make up for some deficit), *affective* (used to lessen inhibiting feelings), and *social* (used to collaborate with others in learning) (pp. 37-57, 135-151). From another perspective, Phakiti (2003) pointed out that learning strategies can be considered traits, since they are part of the long-term memory and are formed with practice, while use strategies are states, as they form part of working memory and so are more associated with the language task being performed.

In an attempt to explore experts' views of learner strategies, Cohen (2005) administered a questionnaire to 23 international scholars concerned with strategy research. A summary of the results suggests that the experts were overall in agreement that learners are conscious when they make their choices of what strategies to use; strategies are used by learners in order or in chorus depending on the nature of the language task they are engaged in; learners' use of strategies relates to such learning styles as being self-autonomous, self-regulated, and self-directed; strategy use is highly

dependent upon the characteristics of learners, the nature of the tasks being performed, and the setting where the performance is taking place; and strategies are used by learners to make their learning and use of language more efficient.

Test-taking strategies.

The kinds of language use strategies employed by learners on the various forms of language assessment are referred to in the literature as test-taking strategies (Cohen, 1998). To be specific, test-taking strategies are techniques that test takers resort to with the aim of getting correct answers on a given test. The successful use of these strategies does not necessarily imply mastery of the testing task at hand, as Cohen (1986) explained clearly when suggesting that test takers may get their answers to a reading task right "without fully or even partially understanding the text" (p. 132). In a later article, Cohen (1992) noted that test-taking strategies represent processes that test takers can have control over by selecting what they believe would help them tackle a test question, suggesting that test-taking strategies are conscious processes. He added that these strategies can either be a short move (e.g., looking for a clue that links the information in the question to that in the reading text) or a long one (e.g., reading the whole text after reading the questions).

Test respondents very often resort to test-taking strategies to compensate for the lack of information that could aid their response to a given question; for example, they may try to match words in the question to ones in the text, simplify the language of the question or the information needed, or use approximation especially when guessing meanings of vocabulary items (Cohen, 1998). However, the use of test-taking strategies is not intended for achieving communicative goals as is the case of language use

strategies; rather, they are used for getting correct answers on language tests. Some scholars count test-taking strategies among learner strategies; for example, Cohen (1994a) mentioned that test-taking strategies are best viewed as "learner strategies applied to the area of assessment" (p. 119). Nevertheless, test-taking strategies are clearly not limited to language learning since the test taking process itself is not limited to the field of language learning, which means that test takers in other assessment settings in other disciplines can certainly use strategies that can aid their performance on a given test in their fields of study.

Scholars (e.g., Allan, 1992; Cohen, 2006; Phakiti, 2008) have referred to two categories of test-taking strategies: *test-management strategies* and *test-wiseness strategies.* The former call for logical and purposeful response behaviors, are reflective of the underlying competence, and are responsive to the underlying construct being assessed. The latter involve the use of textual and/or technical aspects of the test to get the right answers and are not reflective of the underlying competence nor responsive to the underlying construct being assessed. Other scholars (e.g., Cohen & Upton, 2006) have drawn the distinction between strategies and processes such that "*strategies* are subject to control, more intentional, and used to act upon the processes" whereas "*processes* are general, subconscious or unconscious, and more automatic" (p. 2). Similarly, Phakiti (2008) differentiated between strategic processing and automatic processing, suggesting that the more a language user is conscious and aware of the processing underway the more this processing is strategic rather than automatic. By extension, while there are test-taking strategies that are conscious choices (e.g., when a test taker chooses to read the questions first on a multiple-choice reading test before

reading the text), there are test-taking processes that are subconscious operations (e.g., when a high-proficiency test taker attends to the overall meaning using text-level information on a cloze task).

Generally, test-taking strategies are viewed as compensatory since they are typically used by test takers to make up for some deficiency either in the language ability necessary to perform the test tasks, in the skill to take the test, or in both. As such, test-taking strategies can be accounted for in terms of the framework of strategic competence as proposed by Bachman and Palmer (1996). According to this framework, test respondents go through four metacognitive processes when reacting to a testing task: *assessment, goal setting, planning,* and *execution*. That is, test takers first assess the goals of the testing task and determine what aspects of knowledge it draws on. Then, they discern what to do in response to the task. After that, they relate the required information in the task to their knowledge and decide about how to act. And finally, they put what they have decided to do and how to do it into action through the actual provision of the answer. The extent to which learners engage in these processes and manage the use of test-taking strategies on a test can distinguish between performances of two test takers, who might be at the same level of language competency (Bachman and Palmer, 1996).

Cohen and Upton's (2006) rubrics of test-taking strategies (test-management and test-wiseness) on reading tests encapsulate the strategies that were observed in previous research in addition to those strategies they found out through the analysis of verbal reports in their study of test-taking strategies among 32 ESL learners (See Appendix D). Rupp et al. (2006) went further to classify the kinds of test-taking strategies used on reading comprehension tests into general strategies that can be applied to any test format,

text-related strategies that test takers apply to the text, and item-related strategies that test takers use with the question items.

Strategies in reading comprehension.

Reading comprehension has been viewed by reading experts as an interactive and constructive process in which the reader interacts with the text and simultaneously uses a variety of means available to him (e.g., background knowledge and contextual cues) to construct meaning of the text (Dole, Duffy, Roehler, & Pearson, 1991; P. Alexander & Jetton, 2000; Powers & Wilson Leung, 1995). As was the case in language learning, there has been a shift of focus in the context of reading comprehension from the outcomes to the processes involved in reading (N. Anderson, 1991). The study of the strategies used in reading comprehension has offered insights about how readers interact with the text with the aim of understanding it and what resources they draw upon towards this goal (Singhal, 2001).

A distinction has been made between reading skills and reading strategies such that reading skills are used synonymously with abilities to mean traits that readers have developed with practice over time, ranging from lower-level abilities (e.g., word-level processing) to higher-level abilities (e.g., text-level processing), while reading strategies represent conscious processes readers utilize to enhance their understanding of a given reading task (e.g., skipping unknown vocabulary) (Aebersold & Field, 1997; Alderson, 2000; Birch, 2002). Readers tend to exhibit considerable variation in how frequently they use reading strategies depending on what they aim through their reading activities; for example, reading for study demands more frequent use of global strategies (e.g., deciding

what to read closely) and support strategies (e.g., taking notes while reading) than reading for fun (Mokhtari & Reichard, 2008, p. 95).

Previous research (e.g., Garner & Krauss, 1982; Myers & Paris, 1978; Purpura, 1999) has been revealing about the nature of textual processing that highly-proficient readers go through in comparison to low-proficient ones. For example, highly-proficient readers were shown to process textual information at a global level by focusing on the meaning of the text as a whole, using higher-level skills (e.g., understanding main idea, making inferences, synthesizing information) whereas low-proficient readers work with the text at a local level and engage in a decoding process, using lower-level skills (e.g., identifying lexical and syntactic features, making use of various discourse elements). In general, the previous studies of reading strategies with L2 readers at various levels of proficiency in a wide variety of learning contexts point out the crucial role of reading strategies in developing necessary reading skills including comprehension (Alfassi, 2004; Brown & Day, 1983; Mokhtari & Sheorey, 2008; Pressley & Afflerbach, 1995).

The role of metacognition in reading comprehension has been accounted for in terms of metacognitive awareness which involves readers' cognizance of their abilities in relation to the complexity of the reading task, factors pertinent to the difficulty of the reading task (e.g., text familiarity), their repertoire of reading strategies, how the selected strategies are to be used, and the extent to which comprehension is going on (Baker, 2008; Baumann, Jones, & Seifert-Kessell, 1993; Pressley & Gaskins, 2006). This view is consistent with the one that considers "[the] knowledge of text structure … [and the] awareness of the normal sequencing of information in such structures" as two key strategies of reading comprehension (Storey, 1997, p. 221). Metacognitive awareness and

control of reading strategies are crucial to achieve an optimum level of reading comprehension (Mokhtari, Sheorey, & Reichard, 2008). The relationship between metacognition and comprehension can be described as “reciprocal causation” such that “improvements in metacognition contribute to improvements in comprehension, which, in turn, contribute to further improvements in metacognition” (Baker, 2008, p. 34). In this sense, reading comprehension becomes more "constructively responsive” in which process readers adapt their choice and use of strategies to the demands of the textual information they are dealing with (Pressley & Afflerbach, 1995, p. 2). These perspectives have their origin in L1 reading research; for example, in a number of studies (e.g., Baker and Brown, 1984; Cain, Oakhill, & Bryant, 2004; Roeschl-Heils, Schneider, & van Kraayenoord, 2003), it was found that the extent to which readers made use of their metacognitive awareness of their reading abilities and cognitive strategies determined the effectiveness of their reading performance.

In a study of L1 and L2 readers, Sheorey and Mokhtari (2001) found that the level of metacognitive awareness of reading strategies was strongly positively correlated with the level of reading ability and that high-ability readers surpassed their low-ability counterparts in terms of both frequency and efficacy of reading strategy use. Metacognitive awareness of reading strategy choice and use was also found to be a sign of developing the necessary level of comprehension and so was a characteristic of skilled L2 readers (e.g., Barnett, 1988; Sheorey & Mokhtari, 2001). Moreover, research on the differences between good and poor readers in the use of reading strategies has shown that good readers tend to be more aware of what strategies to use and when to use them (e.g., Y. Yang, 2002).

On the basis of the information-processing theory, other scholars (e.g., Phakiti, 2008; Purpura, 1999) upheld the view that strategies can be categorized as cognitive strategies (used to understand and utilize the language) such as *comprehending*, *memory*, and *retrieval*; and, metacognitive strategies (used to act upon and control cognitive strategies) such as *planning*, *monitoring*, and *evaluating*. The use of these strategies during a reading activity can best be described as "a synchronic situation-related variation between cognitive strategies and metacognitive strategies" (Phakiti, 2006, p. 83). Research by Purpura (1998) has confirmed that metacognitive strategies strongly influence and so regulate cognitive strategies during reading. In a more recent study, Phakiti (2006) employed a structural equation modeling approach to analyze the relationships among cognitive strategies, metacognitive strategies, and reading test performance. Phakiti found that memory and retrieval strategies aided comprehending strategies, monitoring strategies regulated memory strategies, evaluating strategies acted upon retrieval strategies, and planning strategies influenced cognitive strategies by means of monitoring and evaluating strategies.

While the role of cognition and metacognition in shaping human learning experiences has been realized through research in psychology since the mid-1970s, only a decade ago did researchers in the field of language testing and assessment show and express an increasing interest in the study of how cognitive and metacognitive strategies relate and influence test performance. One of the first strategy researchers to study the relationship between cognitive and metacognitive strategy use on reading comprehension tests was Purpura (1997, 1998) who observed, through the use of structural equation modeling approach, similar factorial structures for the high- and low-proficiency groups

of test takers, resulting from the two models of cognitive and metacognitive strategy use and the test performance. Nonetheless, the high-proficiency group exhibited more use of such strategies as *inferencing, linking with prior knowledge, practicing naturalistically, self-evaluating* and *monitoring* while the low-proficiency group used more strategies such as *associating, repeating/rehearsing, summarizing, transferring from L1 to L2, applying rules,* and *assessing the situation* (Purpura, 1998, pp. 352-364). Following the same analytical procedure, Purpura (1999) found that the use of cognitive and metacognitive strategies in general correlated with test performance, while memory strategies decelerated it.

Similarly, Phakiti (2003) looked into the differences between successful and less successful test takers in the use of cognitive and metacognitive strategies when responding to a reading test. Phakiti found that the highly successful test takers demonstrated a high degree of awareness of their use of metacognitive strategies in terms of what strategies to use, why they chose these strategies, and how to use them well, which in turn contributed to their performance on the test. He also observed that females and males were almost alike with respect to their cognitive strategy use; however, when interviewed, males expressed more inclination to use metacognitive strategies than females. Moreover, research by Phakiti (2006) and Purpura (1999) indicated that successful test takers used more retrieval than memory strategies, which augmented their use of comprehension strategies. Nevertheless, there is compelling evidence suggesting that highly-proficient readers tended to report fewer instances of strategy use, which can be linked to their advanced ability involving automatic processing that renders strategy use not available for conscious reporting (Purpura, 1998).

With the advent of electronic mediums of textual information, research started to look at the kinds of strategies readers use when interacting with these mediums, and how these strategies compare with those readers use with print materials (Kymes, 2008). Foltz (1996) pointed out that unlike print texts where presentation of information often follows a predictable order, information in online mediums is mostly presented in a non-sequential fashion. According to Foltz (1996), this aspect of online information necessitates adequate level of background or topical knowledge to cope with the insufficient orderliness of online presentation. This is what also makes integration of information, both by relating what is new to what is known and by linking information across multiple sources, an essential skill in online reading (Britt & Gabrys, 2001).

Skilled readers have been shown to benefit from using strategies they have developed and mastered with print materials in their reading of online information. And, they have shown notable tendencies to develop useful reading strategies specific to online environments (e.g., evaluating how credible and trustworthy the information is) (Bland, 1995). In a comparative study of reading strategies used online and in print, Poole and Mokhtari (2008) found that readers tended to use a set of the same strategies in both environments with some variation in terms of frequency, apply certain strategies more often than others in both formats when they found reading to be challenging, and to use online dictionaries with online reading more often than they use print dictionaries when reading print materials.

It is important here to distinguish between test-taking strategies and reading strategies, as these two categories of strategies show some overlap and so can easily be confused in the context of reading assessment. First, test-taking strategies are not specific

to any language skill, although it is true that each language skill has its specific test-taking strategies. However, this does not preclude the fact that some language skills can still share certain test-taking strategies (e.g., a test taker preference to read the questions first on both reading and listening comprehension tests administered in a multiple-choice format). Second, while reading strategies are generally used when readers engage in a reading activity and thus "are related to text comprehension" (Singhal, 2001, p. 1), test-taking strategies are only used when dealing with a test or task which is similar to a test-taking process; that is, the latter are more "driven by the test questions" (Farr, Pritchard, & Smitten, 1990, p. 218).

It is also worth the emphasis here that since there are different types of reading, including reading for test taking, readers tend to utilize different strategies bearing on the goals they try to achieve through the reading activity in which they are engaged (Cordón & Day, 1996). And, because test respondents approach a given reading-test task with the aim of getting it right, they would certainly turn to strategies that involve reading. However, this does not imply that all strategies used on reading tests necessarily relate to the reading process (Allan, 1992), since a considerable number of these strategies are either test-management or test-wiseness strategies or a combination of both. For example, in contrast to non-testing situations, a multiple-choice reading task calls for "a continual, conscious, and linear engagement in problem-solving activities" (Rupp et al., 2006). In practice, Cohen and Upton (2006) observed that their sample of 32 respondents made far more frequent use of test-taking strategies in comparison to reading strategies.

In brief, similar to language use strategies, test-taking strategies are compensatory in nature; however, whereas the former are used for communicative goals, the latter serve

test-taking purposes. The test-taking process of reading comprehension may call upon the use of both test-taking strategies as well as reading strategies; the former are used with the question items whereas the latter are used with the text. In this respect, the use of test-taking strategies far exceeds that of reading strategies since, presumably, test takers' care for answering the question items surpasses their care to fully understand the text. It comes as no surprise that most of the comprehension that test takers attain when responding to a reading test is brought about by their responses to the test questions and not overall understanding of the text. In the section that follows, we will see that test-taking strategies on reading tests are not typically used in an ad hoc way; rather, they are used in systematic and predictable manners that align with test format.

Strategies and test format.

This section revolves about the relationship between task types and formats of reading comprehension tests and test-taking strategy use on these tests.

Aspects of test format determinant of strategy use.

It is intriguing to know what test-taking strategies might work with one test format, but not with another. In fact, previous research has shown that test-taking strategy use on reading tests is substantially determined by the nature of task types when one considers what strategies to use and how to use them (N. Anderson et al., 1991; Nevo, 1989). Nevo (1989) went further to describe the format aspects of a multiple-choice reading test that can influence strategy use significantly, including the level of text familiarity and task complexity. In her study, which sought to examine the use of test-taking strategies on a multiple-choice test of reading comprehension among 42 Hebrew tenth graders studying French, Nevo (1989) noticed that her respondents' ability to use

contributory strategies, or strategies that resulted in correct answers, diminished when they were asked to respond to tests that had unfamiliar texts or difficult question items. For example, the discoursal and pragmatic items, which required test takers to understand the use of the cohesive devices in the texts and so were very challenging, were associated with the use of ineffective strategies on the L2 test. However, this did not apply to all test takers; that is, the use of effective strategies did not always lead to getting the answer correct, which the researcher linked to interfering causes.

Some of the recent studies that looked into task difficulty on reading tests have focused on strategy use among test takers to determine which tasks and question items posed a serious challenge to test takers. For example, Cohen and Upton (2006) observed that the choice of language to report test-taking strategy use among their respondents, while responding to reading tasks on selected practice tests of the TOEFL-iBT, reflected the level of difficulty of the items with which the strategies were associated (for example, one respondent reported that he was *"wrestling with the question intent"* (p. 78) while trying to tackle one of the questions). Other researchers concerned with L1 reading (e.g., Cordón & Day, 1996; Pressley & Afflerbach, 1995) have noted that when readers were provided with an unfamiliar material, they did not only exhibit the use of less effective strategies than that accompanying the use of familiar materials, but also used strategies at low frequencies.

One issue that has triggered extensive research regarding strategy use on reading tests focuses on the extent to which questions on standardized tests are answerable without the texts. For example, Powers and Wilson Leung (1995) had a group of L1 readers answer three sets of reading comprehension questions without the passages, and

at the same time mark on a checklist the kinds of strategies they were using. The findings suggested that respondents' scores were higher than the chance level, but not up to the level that matched the respondents' capabilities. Although respondents had difficulty with some questions that were more passage dependent, they mostly employed their ability of verbal reasoning by using the questions as building blocks to develop a mental schema of the text.

Test-taking strategies dependent on test format.

In studies where student readers were asked to respond to open-ended questions on reading passages (e.g., Cohen & Aphek, 1979), some respondents were observed to try to find where the answer was most likely to be in the passage and write the whole sentence or context containing the answer in response to each given question. In other studies using multiple-choice reading tasks, there was a notable tendency among respondents, specifically those with low proficiency, to guess the key answers from the options without referring to the text, to match the content of the item stem and options with that of the passage (Allan, 1992; Cohen, 1984; Rupp et al., 2006), or to eliminate what they perceived not to be the key answers among the options (Storey, 1997). Farr et al. (1990) was especially insightful in its identification of certain response behaviors on multiple-choice reading tasks. For example, test takers used the questions to guide their dealing with the text in order to identify the key answers. They also used test-taking strategies far more frequently than reading strategies, and showed repeated switching between the questions and the text.

Another study by N. Anderson et al. (1991) examined response strategies with the aim of validating a TOEFL reading test which used three types of questions: main idea,

inferencing, and direct statement. The researchers observed a remarkable consistency among examinees in using patterns of strategies specific to each question type. The researchers noted that question items differed significantly from one another in terms of the kinds of strategies they called for; for example, while inference question items called for more *guessing* and *matching stem with text* strategies, direct statement questions called for examinees' use of *paraphrasing* and *making reference to time allocation* (p. 57). In a more recent study, Rupp et al. (2006) examined response strategies used by a group of 10 ESL readers on three multiple-choice reading tests for academic purposes. Rupp and his colleagues observed that respondents moved from identifying the theme to locating specific details to answer the questions and thus used macro- and micro-level strategies sequentially, used the questions to guide their scanning of the text and locating the key information, used the order of the questions to identify the location of the respective key information in the text, and made use of rational elimination based on prior knowledge or clued-up guessing.

In other studies where cloze tasks were used (e.g., Stemmer, 1991; Storey, 1995), low proficient students were shown to use more micro-level processing when a half of each deletion was given by trying to guess the deleted word using the remnant of it or other local hints. Apparently, these students had a limited ability when it came to understanding the context surrounding the deletion or using macro-level processing. In particular, Stemmer (1991) noted that making inferences was more evident among students taking the cloze test when the number of cohesive devices was low.

Everything considered, research points out that the kinds of formats or tasks that appear on reading tests play a crucial role in determining examinees' strategy choice and

use, and their overall response behaviors. Thus, those strategies that are typically used on cloze formats are different from those used on multiple-choice formats. Examples of the strategies used on the former include *look backward and forward to figure out the missing word* where those used on the latter include *read the questions, then read the text to locate the critical information.* Format aspects such as text familiarity and task difficulty can also impose certain limits on the choice and use of test-taking strategies such that the less familiar the text or more difficult the task, the fewer strategies are expected to be used by test takers. One implication of the dependency of test-taking strategy use on test format relates to how strategies can enable us to check our tests for validity and authenticity, which is to be discussed next.

Strategies and test validation.

The discussion in this section deals with how the study of test-taking strategy use informs the process of validating reading tests.

Premise of validating tests with evidence from strategy use.

Another intriguing facet to the study of test-taking strategies has to do with how it can inform efforts put into test validation. In fact, exploring test-taking strategy use has been viewed among language testers as one of the most important methods of validating language tests (Cohen & Upton, 2006). The process of test validation serves the purpose of confirming construct validity which stands for the extent to which the test measures the underlying psychological construct or concept, or ability, it purports to measure. Early on, the attempts made to validate language tests benefited from the use of associational or correlational measures that were used to correlate the scores examinees

obtained on a certain test with those they had on another, comparable test of the same language skill or ability (Bachman, 1990).

From a strategic standpoint, test validation rests on the question of whether examinees' response behaviors (i.e., test-taking strategies) on a given test conform to the expectations that the test makers have about the test and the purposes for which they designed the test. Language testing experts (e.g., Bachman, 1991) overemphasized the need to attend to any potential disparity between the test constructors' intentions and test takers' perceptions of the test in the design of the test. Consequently, Bachman (1991) suggested that our study of test-taking strategies for the purpose of test validation provides us with a lens into test performance and reflects the extent to which our test tasks assimilate to real-world uses of language. Test validation can be performed during the pilot phase of test development by having a sample from the target population of examinees take the test and observe their test-taking strategies (Cohen & Upton, 2006). The idea of using test-taking strategies to inform test validation was first initiated in L1 testing and has proven useful in both the refinement and standardization of tests (Cohen, 1984).

Using strategies in validation of reading test formats.

The question of what test-taking strategies to use on a given test has to be answered bearing in mind the task or item types that appear on the test (Nevo, 1989). Hence, an incorrect answer to a question item on a test could point out that either the respondent himself failed to answer correctly or the test format influenced the respondent to provide the incorrect answer (Cohen, 1998). Grotjahn (1987), Klein-Braley (1985), and Klein-Braley & Raatz (1984) were among the early researchers who used test-taking

strategies as a basis for validating and refining language tests, specifically cloze tasks. The results of their efforts have led to the development of cloze tests that use rational deletion and offer adequate sampling of the language components to be assessed.

In one study, N. Anderson et al. (1991) attempted to validate a TOEFL reading test in a multiple-choice format by focusing on the relationship between the item types and performances on the test and the test-taking strategies used by their respondents. The researchers found that the question type determined the choice and use of the test-taking strategies to tackle it; for example, wherever respondents were asked to make inferences, some chose to relate and match the content of the question to that of the text. And, wherever a small number of strategies were used, the question items with which such strategies were used were shown to be too easy or too difficult, or less discriminable. Obviously, the number of strategies used with question items can serve to indicate whether these items are adequate as far as their level of difficulty is concerned.

Judgment about the validity of tests is not solely dependent on the behavior of the individual question items on these tests, but on the overall format of these tests as well. In this regard, Tsagari (1994) investigated how the free response format for assessing reading comprehension compares with the multiple-choice format, with a group of ESL learners. On the basis of the examinees' strategy use, Tsagari concluded that the two formats measure reading comprehension differently in that each format calls upon different strategies and so seems to tap into distinct reading abilities. For example, in response to the free response format, students often attempted to locate the key information in the text and use clues to figure out the answer, whereas on the multiple-choice format, students made more use of deductive reasoning and memory.

In a study which aimed to validate a reading test with 13 adult learners of French, Wijgh (1995) observed that test takers' strategy use did not match the test constructors' intentions for the test questions. For example, with question items aimed to have test takers scan the text for superficial information, test takers opted for reading the whole text word by word. This led the researcher to suggest as a potential cause that either the test takers were not skilled enough to use appropriate strategies, or the question items themselves failed to call for test takers' strategy use. Abanomey (2002) sought to explore the effect of text authenticity on the use of test-taking strategies. The researcher did not find an effect of text authenticity on the number of test-taking strategies used by respondents as much as on the nature of these strategies. That is, while the authentic texts invoked the use of bottom-up strategies, the inauthentic texts called upon top-down strategies. The researcher ascribed this observation to the fact that inauthentic texts do not possess the kind of textual features (e.g., cohesive devices) which draw on bottom-up strategies.

With the aim of determining what a cloze task truly measures, Storey (1997) had a group of 25 Chinese respondents answer multiple-choice, discourse cloze tests using rational deletion. The researcher noticed that on tests where the deletions involved discourse markers, respondents were prone to detect the line of argumentation and employ the rhetorical organization to supply the deletions. However, where the deletions were cohesive devices, respondents could just rely on local clues to figure out the deletions. The researcher concluded that discourse cloze has the capacity to call upon processing strategies. Such strategies are used at both local and global levels of the text

and so are reflective of the kinds of processes involved in reading in non-test-taking conditions.

In two of three recent studies, Lumley and Brown (2004, 2006) looked into the validity of the integrated reading and writing tasks on the Next Generation TOEFL (now TOEFL iBT) with 60 respondents from three language backgrounds. On the basis of strategy reports collected from the respondents, the researchers were able to identify serious flaws with these tasks in terms of the difficulty among raters of deciding about whether the responses to the writing tasks were in the participants' words or language they copied from the reading texts. In other words, it was not clear how the respondents arrived at their responses to the writing tasks or even how this related to understanding of the reading texts.

Cohen and Upton (2006) attempted to address the question of the extent to which the reading section of the TOEFL iBT truly assesses the kinds of academic reading skills prospective students need to have command of at the university level. As a result, they noted that test takers dealt with the whole section as demanding of masterful test-taking strategies, and that neither *inferencing* nor the *reading to learn* task types required reading skills distinct from each other. Even so, the researchers concluded that the reading section of the TOEFL iBT adequately measures academic reading skills required at the university level.

In their study of response strategies on multiple-choice reading tests, Rupp et al. (2006) have noticed that whenever test takers were faced with difficult question items, they resorted to logical reasoning on the basis of how a given question item related to the text content; and thus, the test takers were using more lower-order abilities. The

respondents also tended to follow the sequence of the questions to get clues as to where the respective key information is located in the text. According to the researchers, these two observations render the taking of multiple-choice reading tests different from reading for a non-testing purpose in terms of strategy use. All things considered, Rupp and his colleagues concluded that

> [Multiple-choice] questions might function well as separable measures of how difficult different aspects of texts are for test-takers or of how well test-takers engage in lower-order component processes rather than as composite measures of higher-order reading comprehension, which they may be sometimes colloquially assumed to be. (p. 468)

Using strategies to establish construct validity.

Considering that test-taking strategies comprise test-management and test-wiseness techniques, one approach to test validation seeks to ensure that test takers have to rely on the kind of skill or knowledge represented by the test construct to answer the question items more than on test-wiseness. For example, P. Yang (2000) set out to examine the extent to which test-wiseness impacts performance on the TOEFL-CBT. First, the researcher had his respondents answer an adaptation of Rogers and Bateson's (1991) Test-wiseness Test and a TOEFL practice test. Based on their scores on the test-wiseness test, two groups were identified—one as test-wise and the other as test-naïve. Then, respondents were asked to report their strategy use with selected items from both the test-wiseness test and the TOEFL-CBT practice test. It was found that test-wiseness could help with at least half of the items from the listening and the reading sections and

that test-wise examinees could follow systematic ways in tackling those question items amenable to test-wiseness.

In connection with the above findings, Cohen (2006) suggested that test makers should strive to ensure that their tests are not susceptible to test-wiseness if these tests are to be optimally challenging for examinees. In general, test-wiseness strategies fit in the description that Powers and Wilson Leung (1995) offered when stating that "[s]trategies that raise test scores but bear little if any relationship to what the test was designed to measure may diminish the predictive power of a test or dilute the meaning of scores derived from it" (p. 105). This is what also motivated Yamashita (2003) to recommend that test takers' perceptions and opinions about question items be taken into account since test takers are expected to reveal the kinds of strategies that may work for them on given question items, without these strategies being necessarily reflective of the trait or skill being assessed.

Briefly, test validation rests on the question of whether examinees' response strategies on a given test conform to the test maker's expectations or the purposes for which the test was developed. Research has confirmed that the study of test-taking strategy use on reading tests can help us make sure that our tests measure what they are intended to measure, determine how various formats tap into different abilities underlying the main skill of reading comprehension, and ensure that our tests are not susceptible to test-wiseness. This approach has also been found useful in verification of task authenticity; that is, if reading tasks measure the same skills that are part of real-life experiences of reading (e.g., skills of academic reading). One way that can serve as a check of test predictive power is to examine the degree to which high- and low-

proficiency groups of test takers differ from each other with respect to their strategy use. Aspects of differential use of strategies by the two proficiency groups on reading tests form the main focus of the next section.

Strategy use and proficiency level.

In this section, the discussion centers on how test-taking strategy use relates to language proficiency.

Reading and response behaviors of high- and low-proficiency learners.

On the basis of an extensive review of research on reading strategies, Singhal (2001) suggested that there is ample evidence that the use of reading strategies is strongly associated with the level of proficiency such that highly proficient readers use a variety of strategies more frequently and effectively and more knowingly of when and how to use these strategies than less proficient readers. Similarly, Tian (2000) pointed to the major findings from comparative studies of reading strategy use among varied levels of proficiency, which can be summarized as follows: while proficient readers work towards forming a global understanding of what they read using higher-order processing skills, less proficient readers work on a more local level using lower-order processing skills; proficient readers show high flexibility in their strategy use while less proficient readers tend to be more rigid in this respect; and highly proficient readers utilize more active and ongoing monitoring while their less proficient counterparts fail to execute an adequate level of monitoring and so are less able to evaluate and fine-tune their strategy use.

Scholars have begun to look at differences between high- and low-proficiency readers in strategy use on reading comprehension tests. Nevo (1989) was among the first scholars to make the assumption that because reading-test taking represents a problem-

solving situation, high-proficiency test takers employ strategies that increase their chance of getting a question item correct to a greater extent than low-proficiency test takers. She suggested that readers' ability to deal with problem-solving situations in language use correlates with their levels of language proficiency, specifically knowledge of grammar and vocabulary (Nevo, 1989). Also, Purpura (1999) highlighted the role of language ability in the efficient use of metacognitive strategies on reading comprehension tests.

Besides task difficulty, Phakiti (2003) counted proficiency level as another factor that affects cognitive and metacognitive strategy use; for example, he noticed among his respondents that although both high- and low-proficiency test takers exhibited response behaviors so automatic that they were not aware if they had used metacognitive strategies like *checking* and *monitoring,* the low-proficiency test takers employed these two strategies to a lesser degree. However, questions that have yet to be addressed via more research involve whether high- and low-proficiency readers differ from each other in their use of test-management and -wiseness strategies (Cohen & Upton, 2006; Nikolov, 2006) and, if so, how their differential use of these strategies contributes to disparity in their test performance and scores (N. Anderson et al., 1991; Purpura, 1998; Phakiti, 2003).

Test-taking strategies differentiating between proficiency levels.

Early research on the differential use of strategies on reading comprehension tests has pointed out that the main difference between L1 and L2 readers lies in L1 readers' superior ability to make use of intra- and inter- sentential and semantic clues, as measured by means of oral miscue and cloze tasks (e.g., Cziko, 1978, 1980; Douglas, 1981, Hauptman, 1979). This conclusion has been supported by Mangubhai (1991) who

further noted that highly proficient readers exhibit resourcefulness that both serves and automatizes their textual and informational processing to a greater extent compared to low proficient readers.

Mangubhai (1991) based his view on the results of a study he conducted the year before in which he used cloze reading procedures along with think-aloud protocols to investigate strategy use among three different proficiency levels of young ESL learners. The subjects were six EFL learners in year eleven—two high, two middle, and two low achievers according to their scores on a national EFL examination. The subjects demonstrated differences in their strategy use between the high and the low levels of proficiency in that the high achievers used such strategies as *look at larger context after generating the word, refer to prior knowledge, rephrase the sentences in order to generate the word, evaluate guesses for their correctness,* and *analyze the passage using prior and contextual knowledge in order to generate the word,* whereas the low achievers used such strategies as *look at the immediate context and generate randomly and/or reject words on syntactic or semantic grounds* (Mangubhai, 1990, p. 133). Strong positive correlations were observed between the respondents' levels of proficiency, the total percentages of effective strategies they used, and their scores.

In another study combining open-ended and multiple-choice reading comprehension tasks, C. Gordon (1987) used think-aloud protocols to look into response behaviors among 30 tenth-grade EFL learners. The findings from this study suggested that the respondents answered some of the test questions without any indications that they had comprehended the text. Additionally, the respondents who were at a low level of proficiency demonstrated more local processing of the text (i.e., focusing on isolated or

fragmented elements of the text) where highly proficient respondents showed more global processing (i.e., relating the meaning of intact, individual sentences to the whole text). In relation to this, while the low-proficiency students used strategies such as *matching words in the options to words in the text, copying information from the text,* and *translating word for word,* the high-proficiency students used strategies such as *predicting information* and *making inferences* (as cited in Cohen, 1998, p. 100).

Proficiency-related aspects of differential use of strategies.

It is very often the case that in reading test-like situations such as cloze tasks, those who are highly proficient exhibit skillfulness in using text-level comprehension to guide their completion of the cloze items (Bachman, 1985), or at least use other problem-solving strategies such as rational guessing when faced with challenging deletions (Cohen, 1984). On the other hand, those test takers who are less proficient make heavy use of translation and very localized clues to solve deletions (Cohen, 1984). With respect to translation, Upton (1997) and Upton and Lee-Thompson (2001) have concluded that the use of this strategy is inversely related to the level of proficiency and that while L2 readers at different levels of proficiency can benefit from this strategy, low proficiency readers are prone to resort to mental translation in three cases: to deal with unknown words through guessing and substitution using L1 words, to develop understanding at the text level, and to test hypotheses about understanding and to verify these hypotheses.

The tendency to employ effective strategies was observed to be a characteristic of learners' reading in their L1 in contrast to their reading of L2 texts. In her study, Nevo (1989) sought to examine the use of test-taking strategies on a multiple-choice test of reading comprehension among 42 Hebrew tenth graders studying French. The subjects

responded to a multiple-choice reading comprehension test administered first in Hebrew and then in French. While responding to the test items, the respondents were to mark each strategy they used on a strategy checklist on an item-by-item basis. The results showed that the respondents used both effective and ineffective strategies to respond to the reading test items and that more effective strategies and fewer ineffective ones were used when responding to the test in Hebrew than in French. Therefore, the respondents' higher proficiency in L1 when compared to L2 enabled them to use effective strategies or strategies that led them to provide more correct answers on the test in Hebrew (L1) than on the one in French (L2).

Therefore, it seems that the distinction that Dollerup, Glahn, and Rosenberg-Hansen (1982) proposed between the two modes of taking standardized reading tests as "mainline" versus "fragmented" is justified, seeing that mainline reading involves the test taker's skimming of the text to develop the main idea and then answering the question items on this basis, whereas fragmented reading is characterized by, for example, matching of words in the question items with those in the text (p. 96). Therefore, on tests of reading comprehension, while highly proficient test takers are expected to be mainline readers, less proficient ones follow a more fragmented way of taking the test. This may explain why experienced readers were shown to use fewer strategies on standardized reading tests than those who are less experienced, as in Cordón and Day's (1996) study of L1 readers.

Another example comes from a study in which Yamashita (2003) had 12 Japanese EFL students, at the university level, complete a rational deletion cloze test and at the same time verbalize their thoughts. The results showed that the highly skilled respondents

did not spend as much time guessing and inferring the meanings of individual clauses and sentences as did the lower skilled respondents. Rather, the highly proficient readers handled the deletions one after another, drawing on information at both the textual and the clausal levels, whereas their lower proficient counterparts switched back and forth among deletions and used mostly clause-level information.

In an attempt to investigate how strategy use, proficiency level, and level of language aptitude relate one to another, Yoshizawa (2002) had a group of 54 Japanese adult ESL learners respond to a questionnaire. The participants were instructed to report the kinds of text-processing strategies they normally use when performing L2 listening and reading tasks. A language aptitude battery was used to measure the respondents' foreign language aptitude and a test from the TOEFL Institutional Testing Program was used to assess their English proficiency. The researcher found that the respondents exhibited progressively more effective strategy use across their proficiency levels from low to high.

Moreover, the distinction between the two categories of test-taking strategies as test-management versus test-wiseness strategies seems to be pertinent to the differential strategy use of the two discrete levels of proficiency. In this regard, Cohen and Upton (2006) made the observation that their respondents, who were asked to perform TOEFL-iBT reading tasks, used predominantly more test-management strategies than test-wiseness strategies. The researchers linked this tendency to the high proficiency level of the test takers. On this basis, one can deduce that test takers who are at a lower level of proficiency would be more likely to use more test-wiseness than test-management strategies.

In short, as one would expect, there are varying degrees of test-taking strategy choice and use associated with varied levels of proficiency. The previous studies of how test-taking strategies and processes relate to L2 proficiency have revealed that high- and low- proficiency test takers approach reading tests in different ways and exhibit varied response behaviors reflective of their language competency. Moreover, the distinction between the two categories of test-taking strategies as test-management versus test-wiseness strategies seems to be pertinent to the differential strategy use of the two discrete levels of proficiency, in that high-proficiency test takers make more use of test-management strategies than their low-proficiency counterparts whose choices are more limited to test-wiseness strategies. High- and low-proficiency test takers are expected to perform differently on reading tests, which is due in part to their differential use of test-taking strategies. In this regard, high-proficiency test takers are likely to make more effective choice and use of strategies. The following section addresses the question of what aspects determine facilitative versus debilitative effects of strategy use on test performance.

Strategy use and test performance.

In this section, we will see how test performance is shaped by the choice and use of test-taking strategies.

Strategies as a mediator between language competence and test performance.

In their model of communicative language ability, Bachman and Palmer (1996) clearly demonstrated how strategy use, prompted by strategic competence as a mediating component between competence and performance, influences how competence contributes to performance. Whether or not strategy use can make a difference in scores

on language tests used to be an issue calling for research into how the use of test-taking strategies shapes test performance. Meanwhile, test constructors relentlessly expressed their refutation to any claims about the possibility of gaining high scores by means of test-taking strategies (Tian, 2000). Unfortunately, the distinction between the two types of test-taking strategies (i.e., test management and test wiseness) has not been adequately addressed in the debate. Presumably, the test-taking process assimilates to a problem-solving situation, in which case the use of strategies to help deal with this process successfully is natural, and so reflects the authenticity of the test.

Problem-solving strategies on tests of reading comprehension can either be test-management strategies or test-wiseness strategies; hence, while the use of the former is a sign of a skilled response behavior, the latter can be indicative of a poor response behavior or an invalid test item, or both (Allan, 1992). On the other hand, while test-management strategies can be mastered through test preparation practices, test-wiseness strategies are more linked to problem-solving abilities of test takers, and so do not lend themselves easily to training or instruction. In contrast to the use of test-management strategies, as Allan (1992, 1995) suggested, the use of test-wiseness strategies is idiosyncratic, and so results in unfair testing and undeserving achievement, assuming that the given test is amenable to test-wiseness (Cohen, 1992).

Influence of strategy use on test performance.

The issue of how strategy use relates to test performance depends to a large extent on the nature of the test format and tasks. For example, it has been observed across a number of studies of response behaviors on cloze tasks that test takers could still manage to obtain high scores and never had to read the whole text or even understand the main

idea of the text (Cohen, 1984). Test takers in other studies of performance on cloze tasks were observed to first use the local clues to solve as many deletions as they could, and then they moved on to forming a general idea about the text and so used more global clues to help them deal with the unsolved deletions (e.g., Kleiman, Cavalcanti, Terzi, & Ratto, 1986). In another study of strategy use on cloze tasks, Homburg and Spaan (1982) reported that their respondents' use of such strategies as identifying parallel elements, discourse chunking, cataphoric reading, and anaphoric reading correlated with their success in identifying the correct completions. It was also found that those who made effective use of cataphoric reading were better able to figure out the main idea of the text than those who did not.

Cohen (1984) also referred to other studies in which test takers were asked to respond to multiple-choice questions in the absence of the reading passages and how these test takers could still manage to score well above the chance level (i.e., 25% with question items with four alternatives). In his study of strategy use on summarization tasks, Cohen (1994b) found that the time some test takers spent going through and applying strategic processing far exceeded the time they spent writing their summaries. And oftentimes, they chose to add whole blocks of the text being summarized to their responses, which made it difficult to decide about the degree to which these test takers actually used their understanding of the text towards constructing their summaries. There is also considerable evidence regarding positive transfer of strategies from L1 to L2 as far as performance on reading tests is concerned. In this regard, Nevo (1989) identified two strategies as the most frequently used ones in both L1 (Hebrew) and L2 (French) among her respondents, including reading the questions first and then looking for the key

information in the text and matching clues from the question items to those in the text to locate the key information.

As has been confirmed across studies of strategy use on language tests, test-taking strategies are by and large a function of the testing situation and format; in other words, "[a]s long as the task is part of a test, students are bound to use strategies they would not use under non-testing conditions" (Cohen, 1992, p. 99). Bachman and Palmer's (1996) view of the mediating role of strategic competence between language knowledge and language use is obviously warranted in the context of language testing when we consider how strategies can either facilitate or debilitate performance on language tests. For example, a test taker who chooses to read the questions first on a standardized test of reading has a higher chance of completing the test more quickly and efficiently than one who chooses to read the text first and then proceeds to the questions, assuming that both test takers are at the same level of reading proficiency (Cohen, 1992). Clearly, the use of the first strategy can be said to facilitate test performance, whereas the other strategy may considerably debilitate performance under conditions of timed testing and lack of adequate level of proficiency.

Effective versus ineffective strategy use and test performance.

Generally speaking, within the broad field of strategic competence, successful use of strategies demands that strategies be relevant to the nature of the task being performed, strategies be in sync with learner characteristics, and a learner be aware if a strategy is to be used by itself or combined with other strategies and how either form ought to be used (N. Anderson, 2005). Even those strategies that have been shown to be effective can be more or less so in a given testing situation depending on when and how they are used

(Cohen, 1992). Therefore, what Purpura (1998) concluded with regard to how the use of cognitive and metacognitive strategies relates to test performance applies to the case of using test-taking strategies. In other words, the nature of the test task on which a given strategy is used as well as how this strategy is used determine the extent to which this strategy can benefit test performance. Along the same lines, Nikolov (2006) noted that the effective use of test-taking strategies takes into account the degree of compatibility among strategies, the nature of the test task being performed, and command of trends that test takers have developed with the use of these strategies.

The manner in which strategies are used in isolation or in conjunction with other strategies determines their beneficial effect upon test performance and scores (N. Anderson, 1991). N. Anderson (1991) based this conclusion on a study in which he looked at individual differences in the use of reading and test-taking strategies among 28 Spanish-speaking adult ESL learners at three different levels of proficiency. Each respondent was asked to take a standardized test of academic reading, in a multiple-choice format, and simultaneously think aloud his or her response behaviors, either in Spanish or English. Based on case studies of three individual respondents, it was found that the high and low scorers did not differ from each other in the kinds of strategies they used. Rather, the two groups of scorers differed in how effectively they used these strategies individually or in conjunction with other strategies, as well as in their ability to assess and monitor strategy use. The researcher referred to the low scorers' limited repertoire of vocabulary and schema-related knowledge as a potential factor that had constrained their strategy use and, in turn, their performance on the test.

In another study confirmatory of the importance of strategic awareness and monitoring during test taking, Phakiti (2003) examined how the use of cognitive and metacognitive strategies by test takers relates to their performance on an EFL reading comprehension test. The researcher had 384 Thai EFL students, at the university level, take an 85-item reading comprehension achievement test in a multiple-choice format. Then, the respondents reported their strategy use on the test by answering a cognitive-metacognitive questionnaire. The researcher selected the four highest scorers and the four lowest scorers for retrospective interviews. On the basis of both quantitative and qualitative analyses of the data, it was found that the high scorers demonstrated an elevated degree of awareness of their use of metacognitive strategies in terms of what strategies to use, why they chose these strategies, and how to use them well. The researcher concluded that "[i]ndividual test-takers who are metastrategically competent are more likely to understand how the strategies fit together and how they are related to language tasks or TLU [target language use] domains than those with little of this competence" (p. 49).

Test takers of the TOEFL are commonly encouraged to employ certain test-taking strategies which have been shown to be useful. The strategies that can be used on the reading section of the test include familiarizing oneself with the test directions before taking it, reading cursorily and taking mental notes of the closest answers, proportioning the allotted time among the number of question items, using any time left to mark the closest answers to question items whose key answers are not known for certain, and marking C or D options if guessing is not promising (Forster, Karn, Suzuki, & Tateyama, 1997, p. 90). Forster et al. (1997) have also highlighted other test-taking strategies when

advising test takers to read the questions first and then the passage, postpone answering questions about the main idea or the title of the passage until they have answered the other questions, use elimination of alternatives with questions about excluded facts in order to better discern the key answer, rule out obscure and irrational alternatives, and consider more likely key answers those alternatives that are phrased synonymously with or using the same part of speech as that in the key information in the passage (pp. 120-136).

While Forster et al. (1997) were referring specifically to the paper-based format of the TOEFL in giving this account of strategies, their advice is in fact applicable to the computer-based and the internet-based formats of the test. It is obviously the case that on standardized language tests, test takers ought to be familiar with the kinds of test-taking strategies that, as Yien (2001) suggested, can truly mediate between test takers' characteristics including proficiency, and their performance on the test; otherwise, effective strategies would not be characterized as such.

In her study of strategy use on the reading section of the TOEFL-PBT, Tian (2000) worked with a sample of 43 Taiwanese students attending a coaching school. The participants were first asked to take a TOEFL reading practice test and at the same time think aloud their response behaviors. Then, the participants were engaged in a recall task in which they had to write down whatever they could recall from their reading of the test passage. After that, the participants were interviewed as to how they went about preparing for the test and what they thought of the coaching school and the kind of training it offers. A taxonomy of the strategies used was developed, incorporating 42 strategies categorized as technical strategies, reasoning strategies, and self-adjustment

strategies. The examinees were divided into three performance levels: high, middle, and low, as determined by their scores on the test.

The results of the study indicated that the high scorers demonstrated substantial use of strategies that focus on global understanding of the passage with the help of the questions. They used the strategies they were trained to use to supplement the kinds of strategies they have developed themselves, and were notably successful in their adapting and personalizing these strategies. On the other hand, the low scorers worked with the test locally by focusing on individual words and isolated constructions, used the trained strategies as their main strategies, and employed these strategies in the same manner they were trained to use them. The high scorers completed the test in less time and exhibited higher ability in comprehension and information retention as measured by the recall task, compared to the low scorers. In general, the high scorers were observed to use fewer strategies than their low counterparts.

At the level of the three strategy categories, the two performance groups (i.e., high and low scorers) demonstrated variation from each other. In terms of technical strategies, two patterns of variation were apparent. In contrast to the low scorers, the high scorers were found to have developed the ability to start with the passage first and then proceed to the questions and attempt them in the given order, and also the ability to use their understanding of the content of the questions to locate the key information in the text. The two performance groups were different with regard to the level of reasoning they used such that while the low scorers made considerable use of word-based strategies (e.g., matching and association) and micro-level strategies in general, the high scorers were distinctly successful in using macro-level strategies including interpreting and

synthesizing the textual information in an effort to get the gist. The use of self-adjustment strategies was more associated with the use of monitoring and as such was a characteristic of the high scorers. Interestingly, the use of elimination strategy was shown to decrease as the performance level increased across the three levels of performance.

All together, the three performance levels reported in the interviews that they often resort to certain test-management strategies, for example, starting with question items first, answering the main idea question after going through the other items, and using word-based strategies and syntactic or semantic clues when confronted with challenging question items. The results of the interviews also revealed what could be one of the major differences among the three performance levels, namely the awareness of how to use test-taking strategies effectively in terms of what strategies to use, when to use them, and how to use them. Such strategic awareness of effective strategy use increases with proficiency level and so augments test performance, as pointed out earlier in other studies (e.g., N. Anderson, 1991).

Similarly, Nikolov (2006) found that the low scorers among her ESL respondents were more disposed than were high scorers to choose words they were not certain about to complete multiple-matching tasks. This was found through an exploratory study in which Nikolov looked into test-taking strategy use among 12- and 13-year-old EFL learners. A total of 52 participants were randomly sampled so as to represent three levels of proficiency. The data were collected using think-aloud protocols on an individual basis while the respondent was taking a language test comprised of reading and writing tasks. Four case studies were carried out with two top scorers and two bottom scorers, selected on the basis of their scores on the language test. The findings suggested that the two low

scorers processed the text at the word level, where the high scorers showed more tendency to process the text at the sentence level and also benefited, more than did the low scorers, from *relating to self* metacognitive strategy by linking the information in the question item to their real-life experiences in order to check their answers. Overall, although the high performers did not show the use of quite as many test-taking strategies, they made more effective use of the strategies they used, when compared to those with low scores.

In the Cohen and Upton (2006) study reviewed earlier, a major strategy trend among the respondents, whom were rated as highly-proficient, was the use of more test-management as opposed to test-wiseness strategies. Among the most frequently used strategies across the three reading tasks were *read the question then read the passage/portion to look for clues to the answer, either before or while considering options*; *consider the options and postpone consideration of the option*; and *discard option(s) based on vocabulary, sentence, paragraph, or passage overall meaning as well as discourse structure* (pp. 46-102). For the most part, these strategies exhibited clear item-type dependency across the three reading tasks.

To put it briefly, the choice and use of test-taking strategies on reading comprehension tests can either facilitate or debilitate test performance depending on whether or not strategies themselves are compatible with the test format, test takers' level of knowledge and awareness of when and how to use a given strategy, and their level of skillfulness in using a given strategy in isolation or in conjunction with other strategies. Awareness of how to use test-taking strategies effectively in terms of what strategies to

use, when and how to use them can set apart levels of performance and scoring on tests of reading comprehension.

Summary

As mentioned in the beginning of this chapter, in light of the research goals and questions stated in the Introduction chapter, the Review of Literature chapter mainly aims to present a coherent body of the theory and research on factors pertinent to the study of test-taking strategies on reading comprehension tests. The discussion throughout this chapter has demonstrated that the TOEFL iBT, specifically the reading section of the test, remains open to and worth more exploration considering the small number of studies that have targeted certain aspects of this section of the test. The mere fact that the test is evolving makes thoughts about pursuing paralleled lines of research both worthwhile and promising.

There is a convincing body of research that has indicated that the use of test-taking strategies on reading tests far exceeds the use of reading strategies in terms of frequency. Consequently, test takers' care to tackle the question items on a given test surpasses their care to comprehend the text on the test, especially under such constraints as allotted time and low reading proficiency. In view of that, the current study adopted the views that count test-taking strategies to be more associated with the question items than the text (Farr et al., 1990) whereas reading strategies are only those applied to the text (Singhal, 2001). This is postulated to help both avoid any kind of overlap between the two sets of strategies and give a precise account of test-taking strategy use in answer to the three research questions.

This review has indicated that task types and formats on reading tests can largely determine the choice and use of test-taking strategies. The TOEFL-iBT reading section has three different task types, each with distinct item types, in addition to format aspects such as the use of relatively long texts (600-700 words) and three different text types (i.e., argumentative, expository, and historical narrative). This makes it interesting to find out, in answer to the first research question, what test-taking strategies are typically used in response to this section of the test and determine whether or not these strategies differ across task and item types. Cohen and Upton (2006) was the only study that looked for answers to these inquiries, and since the researchers worked with four cultural groups of examinees, they called for attempts to address such research inquiries with more specific populations of examinees. Although the members of the population of this study cannot be said to belong to the same culture, they share the same L1 (Arabic), which suffices to consider dealing with them as a specific population. Another consideration supporting such an attempt, though it was not pursued in this study, is that L1 can exercise an effect on the choice and use of strategies on L2 tests through, for example, positive transfer (cf., Nevo, 1989).

The previous studies of how test-taking strategies relate to L2 proficiency (e.g., C. Gordon , 1987; Mangubhai, 1990, 1991; Yamashita, 2003; Yoshizawa, 2002) have revealed that high- and low- proficient test takers approach reading tests in different ways and exhibit varied response behaviors reflective of their language competency. So far, it has not been clear how these observations apply to performance on the reading section of the TOEFL iBT. In view of the fact that the TOEFL iBT is a measure of language proficiency, test takers at varied levels of proficiency would be expected to perform

differently on the test. Thus, in answer to the second research question, discrete levels of performance are likely to manifest themselves in varying degrees of test-taking strategy choice and use. Notably, this position is in keeping with Bachman and Palmer's (1996) view of strategic competence, which suggests that strategy use plays a mediating role between competence and performance.

The current study did not employ a measure of language proficiency with the participants; therefore, the decision to define the two discrete groups did not derive from their language competence, but from their performance as measured by their scores on the two TOEFL-iBT reading sets. Accordingly, instead of referring to the two discrete groups as high- and low-proficiency test takers, such attributes as high- and low-scorers were deemed more applicable and appropriate. Therefore, through its second research question, this study sought to find out if there are any differences between the high- and low-scoring groups with respect to test-taking strategy choice and use. Since test takers' scores on the reading sets are likely to correlate with their scores in an actual test-taking endeavor, this attempt may tentatively address the question of the role of proficiency level in shaping test takers' decisions about what strategies to use and how to use them.

Previous research (e.g., N. Anderson, 1991; Cohen, 1994b; Nikolov, 2006; Tian, 2000) has also shown that the choice and use of test-taking strategies on reading comprehension tests can either facilitate or debilitate test performance depending on whether or not strategies themselves are compatible with the test format, the level of knowledge and awareness of when and how to use a given strategy, and the level of skillfulness in using a given strategy in isolation or in conjunction with other strategies. Obviously, there are more factors that affect how test-taking strategy choice and use

relate to test performance than can be examined in one research study. Accordingly, in answer to the third research question, the current research limited its scope to aspects that characterize effective strategy use as determined by the success rate associated with individual question items. This applied to those strategies that were used in conjunctions with other strategies, or sequenced strategies. Such aspects as those that can determine efficacy of strategy use included compatibility of strategies with item format and with one another, and the manner in which strategies were used. In spite of the practical values of addressing these inquiries in the context of the reading section of the TOEFL iBT, attempts of the kind of this study are yet to be made.

Chapter III describes the methodology of this research, which includes rationale for the research design, recruitment of participants, instrumentation, data collection procedures, pilot study, data transcription and coding, and procedures for data analysis.

CHAPTER III

METHODOLOGY

This research explored test-taking strategy use among Arab ESL learners when responding to the TOEFL-iBT reading tasks. More specifically, this research attempted to address the following research questions:

1. What test-taking strategies do subjects use when responding to the reading tasks and items?
2. Are there any differences between high- and low-scorers among subjects in their use of test-taking strategies on the reading tasks and items?
3. What aspects of effective test-taking strategy use do subjects tend to employ with the reading tasks and items?

This chapter presents the methodology followed to address the research questions. The chapter is organized into eight sections: (a) rationale for research design, (b) recruitment of participants, (c) instrumentation, (d) data collection procedures, (e) pilot study, (f) data transcription and coding, (g) procedures for data analysis, and (h) a summary of the chapter.

Research Design

This study benefited from an extensive review of the literature on procedures typically used to collect strategy data, such as verbal reports, questionnaires, and checklists. The review pointed to both strengths and weaknesses associated with various procedures especially when one procedure is used by itself. Thus, the current study attempted to avoid major methodological concerns of self-reporting and -revelation procedures mentioned in previous strategy research. On one hand, self-reporting (e.g., questionnaires) provides an indirect way of exploring strategies, and so strategies revealed through this procedure do not accurately represent the actual strategies used in response to specific language tasks (Powers & Wilson Leung, 1995; Purpura, 1998; Wijgh, 1995). On the other hand, self-revelation (e.g., think-alouds) interfere with strategy use on the given cognitive task as well as task performance, which causes strategy reports to either fall short of capturing all strategies used (Afflerbach & Johnston, 1984; Ellis, 2001; Jourdenais, 2001), or be confounded by affective, psychological, and social variables peculiar to the setting where the verbal reports are collected (Mann, 1982).

Therefore, in an attempt to avoid the aforementioned methodological pitfalls, the researcher chose to follow an integrated-method approach through the use of a combination of three procedures (stimulated recall, self-observation, and retrospective interview) shown to be adequate in the collection of strategy data. The decision to use the three procedures together was substantiated by the fact that all three procedures represent retrospective measures; that is, they are used following a cognitive task and they involve eliciting data about how the subject carried out the task. The integration of the three

procedures aimed to get the best of these procedures altogether or capitalize on their strengths, and consequently counterbalance their weak points.

The choice of an integrated-method approach for the design of this study has support in the literature on qualitative research methodology. In this respect, many qualitative research specialists (e.g., Kelle, 2001; Konecki, 2008; Moran-Ellis et al., 2006) have commended the use of a mixed-method approach in order to come to grips with a multi-faceted phenomenon. According to these experts, through the use of mixed methods, the researcher can gain insights into the nature and intricacy of the phenomenon under study, become well-informed about the inner workings of the various elements of a multi-dimensional phenomenon, and increase the likelihood of achieving credible results and findings. In a mixed-method approach, integration of methods can be made by merging these methods such that while each method keeps its own distinct design features, it complements the functions of the other methods and contributes perceptibly and meaningfully to the purpose of the whole procedure (Moran-Ellis et al., 2006). In L2 testing, researchers have generally been encouraged to use a mixed-method approach to get deep understanding of the interaction among constructs reflecting ability and performance, for example, strategy use, proficiency, and test performance (Phakiti, 2003, Yien, 2001).

As stated above, this study combined three methods, namely, stimulated recall, self-observation, and retrospective interview. The three methods represent types of verbal protocols that can reveal "what information [subjects] are attending to while performing their tasks, and by revealing this information, can provide an orderly picture of the exact way in which the tasks are being performed: the strategies employed, the inferences

drawn from information …" (Ericsson & Simon, 1993, p. 220). What follows is a description of these three methods:

Stimulated recall.

Stimulated recall is considered by scholars in the fields of language learning and teaching (e.g., Mackey & Gass, 2005) to be an inner-directed measure in which the learner is provided with a stimulus and engaged in a reflection of the thought processes she had in mind while performing a language task. It is "an information processing approach whereby the use and access to memory structures is enhanced, if not guaranteed, by a prompt that aids the recall of information" (Gass & Mackey, 2000, p. 17). The prompt or stimulus can either be video or audio, or both, and serves to stimulate the learner's recollection of her mental thoughts during the task performance. For instance, having just been in a conversation, learners are asked to listen to an audio-recording of the conversation and use the recording to help them remember their mental processes during the conversation, and at the same time verbalize the kinds of thoughts and reactions they experienced while conversing.

Stimulated recall is used in strategy research for its potential to help researchers "to determine when and if particular cognitive processes, such as search, retrieval or decision making are employed" (Gass & Mackey, 2000, p. 21). One advantage to the use of stimulated recall over think-alouds is that any method-related reactive effects on task performance are avoided (Sanz, 2005). Another advantage is that during a stimulated recall session, the subject is totally focused on recalling and reporting of his thoughts (Kelly, 2009), in contrast to think-alouds where the subject performs the cognitive task and verbalizes his thoughts both at the same time, which Ellis (2001) referred to as "dual

processing" (p. 37). Stimulated recalls help strategy researchers in particular to elicit "task-specific strategy descriptions with corroborating evidence of their use" (Chamot, 2005, p. 58).

Self-observation.

Self-observation can be defined as "the inspection of specific, not generalized, language behavior, either introspectively (i.e., within 20 seconds of the mental event) or retrospectively" (Cohen & Upton, 2006, p. 13). Accordingly, the learner verbalizes the line of thinking he went through while performing a given language task; for example, the leaner may say "[w]hat I just did was to skim through the reading passage for possible surface matches between information in the text and that same information appearing in one of the alternative choices" (Cohen & Upton, 2006, p. 13). Therefore, self-observation can be said to represent a midway solution between self-reporting and self-revelation. The use of self-observation in this study gains support from the work of Pressley and Afflerbach (1995) in which they argued that strategic processes can be verbalized by having respondents take their time attending to more specific bits and pieces of their thoughts and by researchers' providing respondents with prompts to elicit more specific information and explanation (Desimone & Le Floch, 2004; Leighton, 2004).

Similarly, Cohen (1996) noted that strategy researchers are advised to make the shift from using self-report questionnaires to self-observational procedures that enable them to tap into mental processes as close as possible from the moment these processes have been generated. He added that the data collected through self-observation are more likely to "reflect accurately what learners actually do than might the response to a questionnaire item calling for a description of generalized behavior" (p. 13). On the basis

of their research, MacLean and d'Anglejan (1986) argued that the use of self-observation as a retrospective technique furnishes a wealth of information about the kinds of resources learners draw on, be they cognitive or technical, during their performance of language tasks. These researchers also maintained that such information as that obtained through self-observation can serve as an exploratory tool to detect aspects of successful versus poor performance.

Retrospective interview.

Retrospective or post-task interview involves the use of verbal cues to have respondents recollect and report the kinds of mental processes or thoughts they had in their minds while performing a cognitive task, with some elaboration, description, or explanation. A number of scholars have championed the use of retrospective interviews following task performance to elicit descriptions and explanations of thought processes that could help address questions regarding strategy synchronization and compatibility, and the relationships between strategy use and test takers' characteristics (Nikolov, 2006) and test performance (Desimone & Le Floch, 2004; Leighton, 2004 ; Phakiti, 2008; Upton, 1997). In studies where retrospective interviews were used after think-alouds (e.g., Cohen & Upton, 2006; Nikolov, 2006; Upton, 1997), the researchers seemed to agree about the usefulness of retrospective interviews in supplementing and checking data collected by means of think-alouds and facilitating discernment of the kinds of strategies the respondents used.

Other researchers (e.g., Greene & Higgins, 1994; Pressley & Afflerbach, 1995; Yamashita, 2003) have highlighted the value of retrospective interviews, which make use of premeditated prompts, in tapping into processes not readily available for immediate

recall or report. Retrospective interviews can also be useful in eliciting meta-linguistic information that may account for certain aspects of performance (Callies, 2009), steps of decision making during task performance (Robinson, 1992), and motives that spur language behaviors (Ross, 1997). Researchers can use retrospective interviews to check introspective reports for accuracy and to have respondents relate their meta-cognitive and -linguistic knowledge to their reports (K. Taylor & Dionne, 2000).

Retrospective interviews allow the researcher to ask the kinds of questions she wants, paraphrase and elucidate her questions, and use the informant's responses to formulate other questions (Chamot, 2005). A retrospective interview using a stimulus has the capacity to "accurately reveal students' actual learning strategies because it is conducted immediately after a learning task" (Chamot, 2005, p. 113). For example, this can be done by videotaping the learner's performance, playing it back, and pausing wherever the researcher wishes to have the learner explicate his thought processes. Introspective interviews are conducted in a personal fashion. This characteristic of an interview helps the researcher provide the necessary guidance and coaching to the interviewee, attend to so many cues that can provide more information about the interviewee and how he behaved in the research, and collect data that are better reflective of the phenomenon under study than other forms of data collection (Hawley, 2003).

As mentioned above, the integration of the three procedures (henceforth, SRSORI technique where SR=stimulated recall; SO=self-observation; RI=retrospective interview) was an attempt to get the best of these procedures when they are used together, and so offset the weaknesses of each procedure if used by itself. Evidently, there are merits of the SRSORI technique. First, it can produce data comparable to data collected through a

think-aloud procedure and without any interference with task performance. Second, it wards off the problem of boring the subject by having him first perform the cognitive task and think-aloud his mental thoughts, and then sit for a retrospective interview. Such a problem may have serious consequences upon the quality and quantity of the data reported. Third, it has the potential to reveal more information about individual differences in performance of problem-solving tasks than perhaps any other technique used to collect verbal reports. Gass and Mackey (2000) made the same last comment about combining stimulated recall and retrospective interview. Figure 1 below depicts how the use of the three procedures followed a sequence and at the same time how they converged in the production of strategy data in the SRSORI technique.

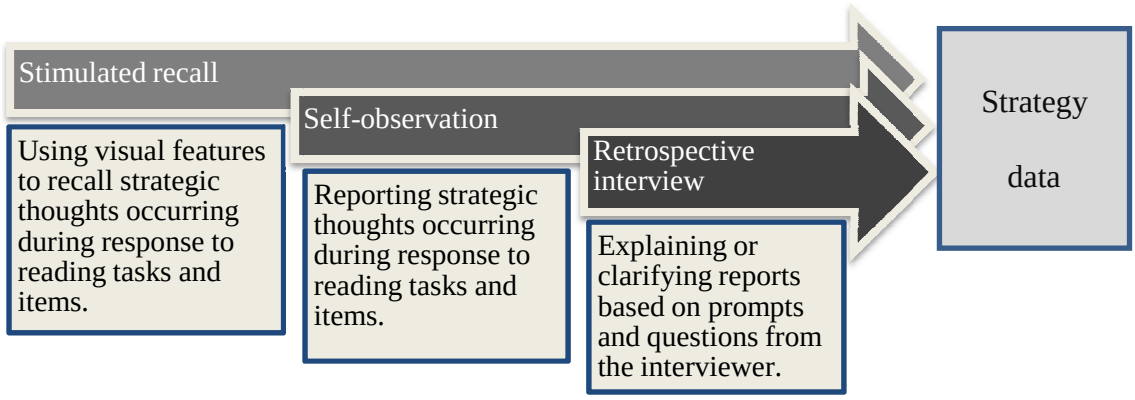

Figure 1. The sequence and convergence of procedures in the SRSORI technique and production of strategy data.

One concern linked with the use of self-observation and retrospective interview is the gradual decline of memory as time passes since the completion of the cognitive task (Cohen, 1986). However, when these procedures are combined with stimulated recall which uses a stimulus or stimuli to augment the informant's recollection of his thoughts during task performance, the problem of memory decline can be countered. Of course,

the time lapse between task performance and verbal protocols, on the other hand, correlates inversely with retention and recollection of mental processes accompanying the performance of the task. Nevertheless, in the case of stimulated recall, the stronger the stimulus is and the sooner the recall occurs after task performance, the greater are the accurateness and completeness of the subject's recollections of the thoughts linked with task performance (Mackey & Gass, 2005; Sanz, 2005).

Besides, researchers who used retrospective interviews (e.g., Polio, Gass, & Chapin, 2006) have called attention to the role stimuli can play in eliciting accurate and complete data. This obviously applies to the use of the SRSORI technique in this study since all procedures combined would be expected to benefit from both the strength of the stimulus and the immediacy of the retrospection process. When using such procedures as those that draw upon memory and retention, researchers are advised to bring into play as many stimuli available during task performance as they can in order to cause the informant "to relive an original situation with great vividness and accuracy" (Bloom, 1954, p. 25).

It was assumed that the use of the SRSORI technique in this study would also benefit from a feature that characterizes self-observation, to be exact, the kinds of reflection and description that the subject makes of thoughts she had during task performance. Another merit is that subjects may reflect on ways of learning and problem-solving specific to their individual cases (Cohen & Hosenfeld, 1981). The kind of stimulus used as part of the stimulated recall process can further the subject's ability to elaborate on and account for her thoughts at every single point of task performance. Retrospective interviews using probes, on the other hand, can allow the researcher to

access the kinds of metacognitive strategies and processes that the informant relied on during task performance (K. Taylor & Dionne, 2000). And, if these interviews are conducted directly after task performance, they can reveal more complete and accurate information on problem-solving techniques and steps (Ericsson & Simon, 1993; Pressely & Afflerbach; 1995).

Towards the end of this section, it is hoped that the aforesaid merits of combining stimulated recall, self-observation, and retrospective interview have rationalized the choice of the SRSORI technique to collect the strategy data of the current study. It is worth noting here that drawing on previous research, precautions were taken against any pitfalls linked with the use of retrospective measures, and recommendations for optimum use of these measures were followed, as is shown below in the section on data collection procedures.

Participants

The sample for the study consisted of 25 Arab ESL learners who were attending undergraduate and graduate programs in the United States at a south-western university at the time the study was conducted. Considering that this study was qualitative in nature and in order to have an adequate number of subjects in high- and low-scoring groups among participants to address the second research question, 19 was assumed to be the minimum number of participants to work with, using normal distribution. Non-probability, purposive sampling was used to recruit human subjects who met certain criteria. First, a subject must have had either the experience of taking the TOEFL iBT officially, or the preparation to take it. Second, the subject must have either taken the TOEFL iBT before 2009, or he or she assured the researcher that he or she had never

responded to or dealt with any of the two reading sets used in this study as expressed by their topics. Taken respectively, the rationale for enforcing each one of these criteria is as follows. First, it was important to recruit participants who were already familiar with the TOEFL-iBT reading tasks to control for the effect of unfamiliarity with the format of the reading tasks, which may disadvantage certain participants. Another important consideration was to ensure that a participant had never responded to any of the two reading sets used in this research; otherwise, certain participants would have been disadvantaged.

Table 2

Participants as per Age Group, Nationality, Gender, Field and Level of Study, and Self-rating of English Proficiency

Age group	Participants	Nationality	Participants	Gender	Participants
(15-19)	4	Egyptian	2	Females	3
(20-24)	12	Iraqi	4	Males	22
(25-29)	6	Kuwaiti	4		
(30-34)	2	Omani	7		
(35-39)	1	Saudi	7		
		Emirati	1		

Field of study	Participants	Level of study	Participants	English proficiency	Participants
Arts	4	Undergraduate	18	Beginning	
Sciences	21	Graduate	7	Intermediate	9
		Masters	6	High Intermediate	12
		Doctoral	1	Advanced	4
Total	25				

Table 2 above displays the participants in the study as per age group, nationality, gender, field and level of academic study, and self-rating of overall English proficiency. (See Appendix F for participant detailed characteristics, including their ages, genders, nationalities, academic levels and majors, self-rated levels of English overall proficiency

and reading proficiency; and Appendix G for participant TOEFL-iBT records and scores on the reading sets used in the study).

Potential participants among Arab ESL learners were offered participation in this research in Arabic, either in person or by phone. Potential participants were offered participation in person in social gatherings that are often organized by the Arab student clubs at the university and the local Muslim community, or by phone after requesting their contact information from their acquaintances who attended these gatherings. Each one of the potential participants was provided with the participant recruitment information, including questions to ensure that the abovementioned criteria were met. Those who expressed their willingness to participate among the potential participants and met the criteria for eligibility to participate in this research were scheduled for interviews at the university library, in an area where the researcher ascertained a quiet and private atmosphere. Overall, participants rated their motivation to perform the TOEFL-iBT reading tasks in this research as above average which is almost equivalent to 4 on a scale of 5. This was determined on the basis of participants' response to the last item on the participant background questionnaire (See Appendix I).

Instrumentation

As mentioned in the Research Design section, the current study made use of a combination of three procedures for collecting qualitative data on strategy use: stimulated recall, self-observation, and retrospective interview. This combination was referred to as the SRSORI technique. In this study, the SRSORI was deemed to be a better alternative to any other procedure for collecting strategy data, including think-aloud protocols. It was also shown earlier how the SRSORI technique combines the strengths of the three

componential procedures, and so counteracts the weaknesses associated with each procedure if used by itself. The use of this technique followed the participant's response to each of two computerized-reading sets in two sessions, one at a time. The participant was asked to make use of a stimulus, in the form of a playback of screen recording of his or her response to each reading set, to help him or her recall the kinds of thoughts he or she had in mind while responding to the reading set. He or she was instructed to describe and explain these thoughts as much as he or she could. At the same time, the researcher provided necessary prompts and asked opportune questions to elicit more details. The two SRSORI sessions for each participant were mainly in Arabic, and were both audio-recorded in a digital format.

Thus, before the respondent was engaged in the SRSORI procedure in each one of the two data collection sessions, he or she responded to a reading set. That is, there were two reading sets having the same tasks that appear on the actual TOEFL-iBT reading section. The two reading sets appear as part of one of two authentic tests in *The Official Guide to the TOEFL Test* (ETS, 2009). The guide has a companion CD that has the two tests administered in a computerized form and so it uses automated scoring with the objective sections of the tests, including the reading section. Both the administration and scoring of the reading sets on the CD are analogous to those of the actual TOEFL iBT. Each reading set consisted of a 600-700 word text followed by 12-13 question items. One reading set was chosen to be from a field that belongs to the arts, whereas the other was from a science field. This step aimed to balance out the effects of text content of the reading sets across two potential groups of participants according to their fields of study. The use of two reading sets with participants in this study was considered appropriate

since the reading tasks and items on the TOEFL were designed such that the whole reading section possesses a high discriminatory power among upper levels of English reading proficiency (Enright et al., 2000). This is especially the case when one considers that 22 participants, almost 90%, rated their reading proficiency in English as either Intermediate or High Intermediate.

Each reading set includes test items designed to evaluate reading for Basic Comprehension, Inferencing, and Reading to Learn skills. Basic Comprehension (BC) has five item types: Vocabulary (BC-*v*), factual information (BC-*fi*), negative fact (BC-*nf*), pronoun reference (BC-*pr*), and sentence simplification (BC-*ss*). The Inferencing (I) three item types are basic inference (I-*bi*), insert text (I-*it*), and rhetorical purpose (I-*rp*). The Reading to Learn (R2L) tasks are prose summary (R2L-*ps*) and schematic table (R2L-*st*). Table 3 below outlines descriptive information about the reading sets used in this study, including the topic of each set, type of text genre, total number of question items, item types, and total score. Each item is worth one point except for the last item on both reading sets, which receives partial credit (0-2 points for item 13 on reading set 1 and 0-3 points for item 12 on reading set 2).

As stated by the ETS (2009), the two tests in the TOEFL-iBT guide and on the CD companion are actual TOEFL-iBT tests used in previous administrations. The ETS (2009) argues that the two tests can provide the test taker with "an estimate of how [he or she] would perform on the actual exam" (p. 3). The TOEFL iBT has established validity and reliability (See also ETS, 2011a, 2008, 2011b), and uses unbiased objective scoring (ETS, 2009). As for authenticity and face validity (cf., Alderson, 2005), the kinds of testing tasks used on the TOEFL iBT mirror the tasks students are expected to perform in

academic settings (ETS, 2009). (The reader is referred to Chapter 2, Section 2 where issues pertinent to validity and reliability of the TOEFL are spotlighted; and Section 3 where task formats and characteristics on the TOEFL reading section are described in detail).

Table 3
Descriptive Information about the Reading Sets

	Topic of set	Type of genre/field	Number of items	Item types	Total score
Reading set#1	Nineteenth-Century Politics in the United States	historical, arts	13	1. BC-*v* 2. BC-*fi* 3. I-*rp* 4. BC-*fi* 5. BC-*fi* 6. BC-*v* 7. BC-*fi* 8. BC-*v* 9. I-*bi* 10. BC-*nf* 11. BC-*ss* 12. I-*it* 13. R2L-*ps*	14
Reading set#2	Geology and Landscape	expository, sciences	12	1. BC-*fi* 2. BC-*v* 3. I-*bi* 4. BC-*v* 5. BC-*fi* 6. I-*rp* 7. BC-*v* 8. BC-*pr* 9. BC-*ss* 10. BC-*fi* 11. I-*it* 12. R2L-*st*	14

Accordingly, in order to elicit participants' reports of the kinds of test-taking strategies they tend to use on the TOEFL-iBT reading tasks, each participant was first

asked to respond to reading sets in two sessions. Then, the participant was engaged in a SRSORI procedure to have him or her report the kinds of strategic thoughts and behaviors he or she made use of in response to the tasks and items on the reading sets. The scores the participants received on the reading sets were used to determine which participants fit in the high- and low-scoring groups. Further, based on score reports, participants' successful answers to the question items on the reading sets were used to identify effective aspects of test-taking strategy use in relation to task and item types.

Data Collection Procedures

Data collection was started in mid May, 2010 and was completed by mid August of the same year. Before collecting data for the main study, a pilot study was conducted to check the adequacy of, and make any necessary refinements in, research procedures and materials to be used in the main study. Arabic was used as a medium of communication in all data collection procedures in order to ensure that all participants understood what they were expected to do and to collect data not affected by any language deficiencies in using English. Nonetheless, a mix of Arabic and English was permitted so long as this did not interfere with providing accurate and complete reports of strategy use on the reading tasks. (The pilot phase of this research is described below in detail, after introducing data collection procedures followed in the main study).

Before data collection.

The materials that were crucial elements of the data collection process were made available and ready, including a laptop computer, the TOEFL-iBT guide CD companion, the screen recorder program, and a SONY digital recorder. The screen recorder program, named BB FlashBack Express (Blueberry Software, 2009), records any online activity, or

activity that appears on a computer screen, and produces a movie of the whole activity. This program was used in this research as a stimulus, and proved to be a significant feature of the SRSORI technique, as will be shown in the section on the pilot study. Since this research called for participation of human subjects, it was mandatory to obtain the approval of the Institutional Review Board (IRB) at the university. The IRB approval implies commitment on the researcher's part to give priority to the welfare of participants at all various stages of the research. After receiving the IRB approval, almost 90% of the potential participants were recruited to participate in the study, in accordance with the IRB guidelines for recruitment of human subjects. The remaining 10% of the potential participants seemed either unable or unwilling to participate. Interviews were scheduled with participants to be conducted at a time convenient to them.

The data collection setting was selected such that it provided an atmosphere as quiet and conducive to optimal test-taking and interviewing tasks as possible. The researcher's first priority was to establish rapport with the participant through informal conversation in order to facilitate subsequent communication. The researcher used a prepared step-by-step guide as to what do from the start to the end of each data collection session. The participant and the researcher sat side-by-side at a table on which the laptop computer, the audio-recorder, and the other materials were placed. As mandated by the IRB, the participant was first requested to read the consent form and sign it upon agreeing to participate in the research.

The IRB consent form described the purpose of the research, the procedures to be followed, how a participant's data would be secured and kept private, the form of compensation, and contact information of the people concerned with the research. After

signing the consent form, the participant was requested to maintain as confidential any information about his or her participation in this research, including materials and procedures used in data collection, in order not to cue other potential participants in any aspect of the research. (Appendix H has the research approval letter from the IRB, and other approved documentation including the information used to recruit participants as well as the informed consent document both in English and Arabic).

After completing the setup of the research materials, the participant was requested to complete a participant background questionnaire (See Appendix I). Next, the participant was oriented to the research procedures and steps. The participant was requested to deal with the reading sets in the same manner he or she would with actual reading sets on the TOEFL iBT. In order to motivate participants to respond to the reading sets to the best of their abilities, to try to achieve the highest score as they could, and to make this experience of consequential value to them, each participant was informed that he or she would be rewarded a sum of money from $10 to $20 depending on his or her total score on the reading sets.

Then, the participant was introduced to verbal report, and given directions on how to verbalize his or her thoughts that accompanied the performance of a given task. The participant was directed that when producing his or her verbal reports, he or she should express any thoughts he or she had in mind while responding to the reading tasks. The participant was informed that he or she could use Arabic, or a mix of Arabic and English only if this did not interfere with his or her providing accurate and complete verbal reports. The participant was also informed that the researcher would provide some

prompts or ask questions to elicit more information or clarification from him or her as part of the verbal report.

Next, verbal report was modeled by the researcher using a grammar exercise instead of a reading task so as not to lead or influence the participant towards reporting certain reading-test-taking strategies, or to ward off the so-called halo effect. The participant was then asked to practice verbal report, using a reading mini-test composed of a short text followed by a multiple-choice question item. Before practice, the participant was reminded that he or she was expected to express any thoughts he or she had in mind while responding to the reading mini-test, using Arabic or a mix of Arabic and English as he or she wished, and that the researcher might provide some prompts or ask questions to elicit more information. The practice step was timed in order for the participant to work under roughly the same conditions he or she would experience when responding to the reading sets. Once the participant signaled to the researcher that he or she had decided on an answer to the question, he or she was requested to verbalize whatever he or she thought about to get to the answer to the question. The researcher provided the participant with prompts and questions whenever further clarification or explanation was needed.

At the end of the practice session, the participant was provided with feedback on his or her verbal report, and offered recommendations as needed. The training was also aimed to familiarize the participant with the researcher and the setting. Assuming that the reading mini-test could serve as a stimulus, it can be noted that each participant was also trained in stimulated recall (Mackey & Gass, 2005). At the end of the training session, the participant was asked if he or she had any questions about verbal report, and his or her

questions, if any, were answered. Arabic was used during the training session at the three stages of giving instructions, modeling, and practice. (See Appendix J for procedures of verbal report training in English).

Afterwards, the researcher ran a sample reading set (entitled *The Expression of Emotions*) to demonstrate for the participant how to go through the reading tasks, starting from the instructions to pressing the Finished Button to end the test-taking session. The tutorial involved going through the instructions; displaying how to use the interactive features of the interface, including a demonstration of how to move to the next question and how to go back to a previous one, how to go to the passage in a full view, and how to select answers to the question items; and demonstrating what to do after completing the reading set. It was brought to the respondent's attention that he or she should try to complete the reading set before the time expires.

Although participants were already familiar with the formats of the task and item types on the test, it was assumed that providing them with such a tutorial would help in two ways. First, it would refresh their memories of task and item formats. And second, it would minimize the effect of any differences among them that are due to time length since they last took the test. Using a short clip of the screen recording of the sample reading set, a demonstration was provided of how to use the arrow on the screen to express thoughts linked with certain movements of the cursor during the response activity. These thoughts were basically about technical matters (e.g., run the program, press Continue Button, choose an answer, etc.) in order not to cue the participant in certain test-taking strategies. The participant was informed that he or she would see a screen recording of his or her response to each reading set, and that he or she would use

this recording as a stimulus to help him or her recall the thoughts he or she had in mind during test-taking. At the end of the tutorial, the participant was asked if he or she had any questions and if he or she was ready to respond to the first reading set.

During data collection.

Once the participant expressed his or her readiness to respond to the first reading set, the researcher ran the screen recorder, and immediately the participant was instructed to start responding to Reading Set 1. The reading set was timed for 20 minutes and the time remaining showed on the window the participant was interacting with. The time allotted for the participant to complete a reading set was similar to that of the actual TOEFL iBT. Consequently, this presumably made the participants live the experience of actual test-taking, and so their strategy use in the research setting would be considerably similar to that in an actual TOEFL-iBT setting (N. Anderson et al., 1991). Next, the researcher checked to make sure that both the TOEFL-iBT program and the screen recorder were running well, and then he left the participant to work on his or her own and sat at another nearby table, yet not directly visible to the participant. Upon the participant's completion of the reading set, the researcher stopped the screen recorder and saved the recording as a video file, using a code composed of the participant's ID number and either A for Reading Set 1 or B for Reading Set 2.

A few participants were not able to complete one or the other of the given reading sets before the time expired; thus, these participants were permitted to complete the reading set even after the time expiration. Otherwise, ending the testing session by time expiration would have led to excluding the last item type, namely the R2L task item, from the collection and analysis of data from these participants. In all of these cases, the

participants were able to press the Finished button in just one to two minutes after time expiration. Because the TOEFL-iBT guide software is intended to serve practice purposes, it allows the user to work on reading sets even after the time expires.

Upon the respondent's completion of the reading set, the researcher informed him or her that the verbal report procedure would be started. The researcher operated the audio-recorder and played the screen recording. Then, the participant was engaged in stimulated recall combined with self-observation using the screen recording as a stimulus, while the researcher was listening attentively, providing the participant with specific and purposeful prompts and/or questions to elicit further clarification or explanation. Such prompts and questions made use of visual cues to enable the participant to remember many more details. To avoid interrupting the respondent's recall and self-observation, prompts were generally provided whenever the respondent paused. Also, the questions were often posed when the respondent showed that he or she had made a complete verbalization of thoughts linked with the response to a given question item.

Guidance was also offered when it was needed, for example, to have the participant verbalize his or her mental thoughts as they occurred during the test-taking activity when he or she, for example, started describing what he or she usually does in response to a given task or item. To control for the effect of individual differences among respondents in the manner in which they reported their strategy use, the researcher offered prompts and questions to respondents in more of an informal, relaxed, and individualized manner in order to ensure more effective and complete reporting on the respondent's part (Patton, 2004). Next, the respondent was offered a break in between two data collection sessions. Afterwards, the participant was asked if he or she was ready

to respond to the second reading set. Upon the participant's agreement, the researcher ran the screen recorder and immediately instructed the participant to start Reading Set 2. Then, the same procedures as those followed in the first data collection session were followed exactly in this second session.

After data collection.

At the end of the whole data collection procedure, the participant was provided with his or her scores on the reading sets, using a feature in the TOEFL-iBT guide CD that shows scores and correct versus incorrect items. Besides, this feature furnishes the user with key answers along with explanations for these answers. Based on his or her total scores on the two reading sets, each participant was rewarded a sum of money from $10 to $20. The total score was 28. The subject who obtained a total score from 26 to 28 received $20. A score from 21 to 25 received $17. A score from 16 to 20 received $15. A score from 11 to 15 received $12. A score from 0 to 10 received $10. The participant was then debriefed about his or her performance and thanked for participating in the research.

Pilot Study

In order to check the suitability of the SRSORI technique and the reading sets, as well as the procedural steps to be followed in the main study, a pilot study was conducted with five volunteers from the target population of the study. Based on the pilot study, the instructions and directions that were part of the orientation and training of participants were made clearer and more elaborate. The modeling and the training steps of verbal report were improved in two ways. First, the modeling was made expressive of more details so as to make it clear to the participant that while producing verbal reports, he or she is expected to provide as much detailed reports as possible. Such reports should

include even mental and physical behaviors that participants may consider unworthy of mentioning, because these behaviors might still be valuable to the research. Second, the practice step was timed in order to simulate the experience of responding to the reading sets under time allotment. Another reason was that practicing with no time limit might cause participants to use strategies they would not use in real-testing conditions.

Table 4 below displays volunteers for the pilot study as per age group, nationality, gender, field and level of study, and self-rating of English proficiency. Data collected in the pilot study were not analyzed, nor used towards the attainment of the results and findings of the main study.

Table 4
Volunteers for Pilot Study

Age group	Participants	Nationality	Participants	Gender	Participants
(20-24)	1	Egyptian	2	Females	2
(25-29)	3	Iraqi	1	Males	3
(30-34)	1	Libya	1		
		Saudi	1		

Field of study	Participants	Level of study	Participants	English proficiency	Participants
Arts	3	Undergraduate	1	Beginning	
Sciences	2	Graduate	4	Intermediate	
		Masters	2	High intermediate	2
		Doctoral	2	Advanced	3
Total	5				

The use of the screen recording feature as a stimulus in the SRSORI procedure gained support from all of the volunteers in the pilot study. (For illustration, Figure 2 below shows a snapshot of the sample reading set in the display mode of the screen recording program). At the end of their interviews, the volunteers were requested to answer the following questions that were presented to them in Arabic: 1. *How do you evaluate your recall of your thoughts during verbal report?* (Circle a percent: 10% - 20%

- 30% - 40% - 50% - 60% - 70% - 80% - 90% - 100%), 2. *How useful did you find the screen recording to be in helping you recall your thoughts during verbal report? Please include examples of these thoughts. Please note: Your feedback will help the researcher make a wise decision as to whether to use this feature in his research or think of an alternative.*

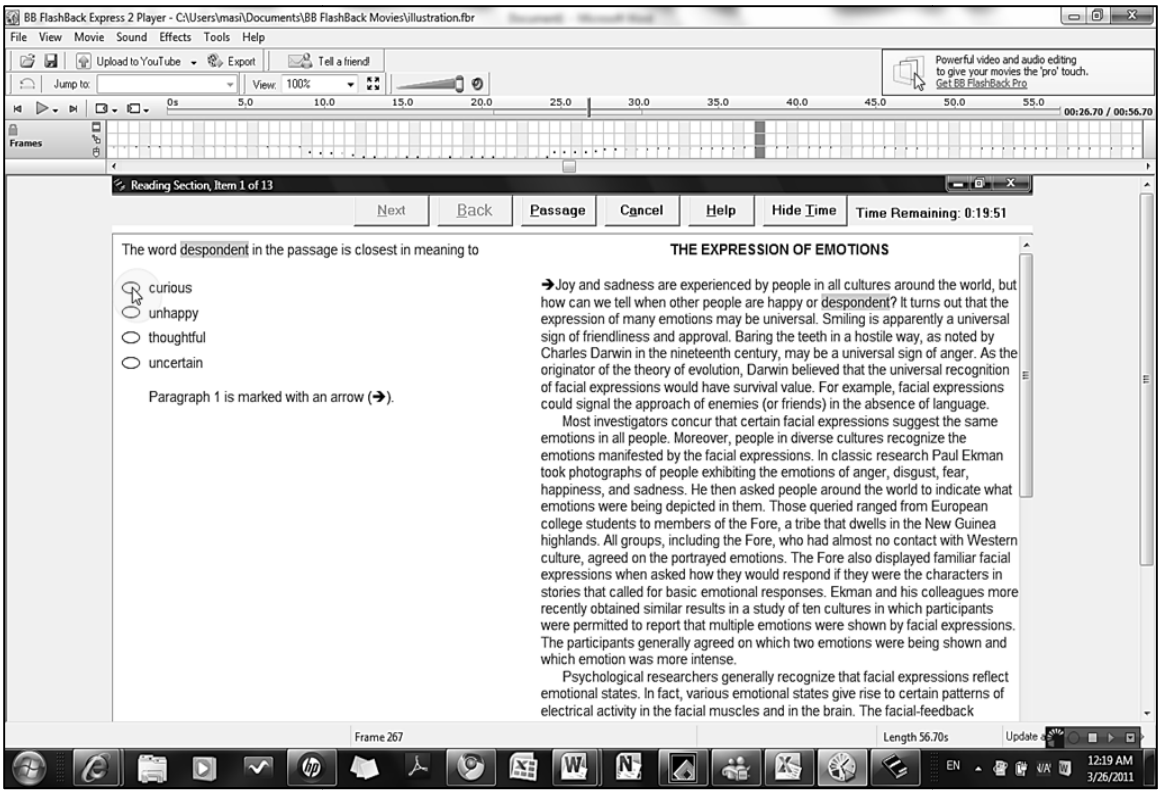

Figure 2. This is a snapshot of the screen recording of the sample reading set. It clearly shows all visual details in the actual online activity. The cursor also appears highlighted.

On average, the volunteers indicated that the playback of the screen recording during the SRSORI procedure helped them recall at least 80% of the thoughts they had in mind while responding to the reading sets. Here are translations of the comments the volunteers made about the screen recording. The comments are preceded by the academic degree the volunteer was seeking and the percentage he or she assigned to the usefulness of screen recording:

- MS (90%) “The screen recording feature helped me in a few cases when I did not recall my choice or how I reached that answer. Following the cursor was a good clue.”
- BS (80%) “I would not have been able to remember so many details without the recorded clip.”
- PhD (80%) “It was very helpful in that I could track my thoughts back and see how I was able to make decisions regarding the choice of the most appropriate answers.”
- PhD (70%) “The program helped me remember my ways of thinking, including what I was doing with certain question items and distracters during the stop time and at moments of hesitation.”
- MS (80%) “The program helped me recall my thoughts through the use of the computer interface and the cursor motion between the questions and the passage. I could recall points of hesitation through the cursor motion and what I was thinking about during stop moments.”

Other improvements made on the basis of the pilot study included providing participants with a tutorial on how to respond to the reading set using a sample set having the same format as those sets used in the actual data collection sessions. The tutorial was necessary because some of the volunteers in the pilot phase of this research expressed their forgetfulness of what some of the task items looked and how to select the answer to certain task items. It was also seen necessary to provide a demonstration of how to use the cursor in the sample clip to express thoughts linked with its movements during the response activity. In addition, the pilot phase helped the researcher develop the habit of providing selective prompts during the SRSORI procedure. These improvements can be

said to have enhanced both validity and reliability of data collection materials and procedures.

Procedures for Data Transcription and Coding

Once data collection was completed, the procedures for data analysis were started. Because the data collected in this study were basically verbal reports, the researcher followed both quantitative and qualitative approaches to data analysis. In order to make the data ready for analysis, the data had to be transcribed and coded. And because most of the data were collected in Arabic, they had to be translated into English first. This process was made easier by the fact that some respondents produced most of their verbalizations, or at least some of them, in English, either by producing whole expressions or using key terms and phrases. What follows is a description of the transcription and coding procedures that were followed in this study.

Using a vocabulary item as an example, here is an explanation of the terminology used in data analysis and description:

Example:

The word perplexing in the passage is closest in meaning to

- difficult
- ambiguous
- confusing
- complex

The first part of the item that presents the question is called the “stem”. The list of the four words that follow the stem are called the “options”. Thus, the stem and the options make up the item. Because Option 3 is the correct answer, it is called the “key”, while the other options are referred to as “distractors”. Another term to be used very often is

"critical information", which is the portion of the passage or paragraph where information pertinent to the answer to the item can be located. (Refer to Appendix C for descriptions of the TOEFL-iBT reading task and item types as well as their respective terminology).

Task analysis.

Test takers responding to the reading tasks on the TOEFL iBT may choose to follow certain routes, among many alternatives, to provide the answer. For example, they may follow an idealistic way of tackling a given task item. That is, they read the question first and then the given paragraph, read all of the options or answer choices available, and decide which one can best fit as an answer to the item. Or, they may choose to read the question first and then the options, read the paragraph, and reread the options to choose one of them. Indeed, these two response behaviors are the most common ones, but this certainly does not preclude the fact that other response behaviors are still possibilities.

Transcription.

Bearing on task analysis and relevant literature (e.g., A. Green, 1998; Gass & Mackey, 2000; Mackey & Gass, 2005), it was assumed that the best approach to transcribing and coding the verbal data would mainly be made in relation to the research goals and questions. Another consideration that influenced how the verbal reports in this study were transcribed was the fact that verbal reports themselves were strategy data. In this study, strategies represent problem-solving tactics used by the test taker to help answer a given test item to the best of the test taker's ability and use of available resources. Therefore, strategy data were transcribed in the form of sequences of strategic moves (Ericsson & Simon, 1993; Gass & Mackey, 2000) — each represented a unit for

analysis in subsequent stages. The strategic moves used with each task item were aggregated into episodes, and each episode was marked by a timestamp indicating the time at which the verbalization of the response behavior was recorded (A. Green, 1998; Ericsson & Simon, 1993). In this kind of analysis, *segment* names a unit constituent of the episode, and so it can be used to signal a strategic move.

All verbal data that expressed strategic moves were transcribed. Nonverbal forms of language and pause time were not considered in the transcription process as there was more than enough data without them. Screen recordings were consulted to resolve a couple of cases of ambiguous or incomplete verbalizations. Transcripts were concurrently checked for both accuracy and completeness by playing back their respective audio-recordings (A. Green, 1998). The researcher's questions were not transcribed for they were all intended to elicit further description from participants of their response behaviors or clarification of their verbalizations. (See Appendix K for a sample transcript).

Item sampling for coding.

After verbal data were transcribed, they were prepared to be coded. Since there are more vocabulary and factual information items in the reading sets than the other item types in total, and because of the huge amount of the verbal data collected in this study (an average of 50 minutes of recording time per participant), it was felt necessary to select for coding the verbal reports of responses to only certain items. The decisions about which verbal reports to code and consider in subsequent procedures of data analysis were made based on item analyses. Therefore, responses to test items which were to be coded and analyzed were sampled from each reading set on the basis of two criteria. First, these items adequately represent the various task and item types that appear

on the reading section of the TOEFL iBT, as expressed earlier in the design and format aspects of the TOEFL reading section in Chapter 2. Second, the items have acceptable, or nearest acceptable, values on item difficulty (*p*) and discriminability indices (r_{pbi}).

Item difficulty indicates the proportion of test takers who responded to the item correctly (range: $0<=p<=1$), whereas item discriminability refers to the difference between the proportion of the upper scorers who got the item right and the proportion of the lower scorers who got the item right (range: $-1<= r_{pbi}<=1$). The higher the *p* value for an item, the easier the item is. And, the higher the r_{pbi} value for an item, the more discriminable this item is. With this in mind, items were selected such that they are of average difficulty ($0.33<=p<=0.66$) and reasonable discriminability ($r_{pbi}>0.25$). This step was expected to serve data analysis in two ways. First, responses to items within these ranges of *p* and r_{pbi} would be those that had more strategic content compared to other items that fell outside these ranges. And second, responses to items in these acceptable ranges would involve distinct aspects of strategy use between scoring levels among participants.

Drawing on scoring and performance data of participants, Cronbach's Alpha (α) for the two reading sets as a whole was found to equal 0.87 which is considered very good in educational measurement and research. Cronbach's Alpha is a measure of the internal consistency or reliability of items on psychometric tests. Internal consistency reflects the strength of the correlations among the test items and their collective ability to assess a given construct. A coefficient alpha of 0.87 suggests that the two reading sets are considerably consistent in measuring reading proficiency across their items. The estimates of Cronbach's Alpha, item difficulty, and item discriminability were arrived at

through the use of Iteman 4, a software program designed by Assessment Systems Corporation (2010).

However, some items were selected though they had values either too high or low on the *p* index or too low on the r_{pbi} index. These items were selected only when shown to have the closest values to those most desirable among the items of their types on both of *p* and r_{pbi} indices, or on either index when the item was equal, or near equal, to another item on the other index. When there were two or more instances of the item on both of the reading sets (inc., BC-*v* and BC-*fi* items), two instances of the item were selected such that each instance of the item appears on either reading set. This was done in keeping with the desirable values on *p* and r_{pbi} indices as mentioned above.

The reason why instances of certain item types were selected across the two reading sets was because participants were expected to perform differently on these items, depending on their field of study (arts vs. sciences) (cf., Xi, 2010). There were yet sole item types, or item types with single appearance on the reading sets (incl., BC-*nf*, BC-*pr*, R2L-*ps* and R2L-*st*), which demanded that a single instance of each one of these item types be selected. (See Appendix L for item difficulty and discriminability values, Cronbach's Alpha (α) if the item is deleted, and the decision about whether or not to include the item and the verbal report associated with it in data analysis).

Another reason that called for the use of item analyses to decide which verbal reports to code and analyze was motivated by research on problem-solving tasks, which suggests that tasks that are reasonably difficult can call for more strategy use than do tasks that are either easy or difficult (Ericsson & Simon, 1993). Also, one interesting feature of using tasks that would engage more strategy use is that these strategies can be

brought to a level of consciousness enough to allow for adequate verbalization or description of these strategies (Mangubhai, 1990). In the case of the current study, the use of the reading sets served these two purposes well in that all participants engaged in strategy use regardless of their levels of self-rated proficiency or scoring. Second, all participants had no noticeable problem reporting their strategy use during the SRSORI sessions, which indicates that they reached a desirable degree of consciousness of strategy use. Furthermore, items with high values on the item discriminability index specifically were expected to require distinct aspects of strategy choice and use by the higher versus lower scorers. Table 5 below shows the distribution of item types whose verbal reports were coded and analyzed for each participant.

Table 5
Distribution of Items whose Verbal Reports were Coded and Analyzed per Participant

Item type	Item # on Reading Set 1	Item # on Reading Set 2	Total items per item type
BC-*v*	1	4	2
BC-*fi*	7	5	2
BC-*nf*	10		1
BC-*pr*		8	1
BC-*ss*	11	9	2
I-*bi*	9	3	2
I-*it*	12	11	2
I-*rp*	3	6	2
R2L-*ps*	13		1
R2L-*st*		12	1
Total items	8	8	16

Note. There are single items for item types BC-*nf*, BC-*pr*, R2L-*ps*, and R2L-*st* across the two reading sets.

Coding.

Once the verbal reports to be coded were decided upon based on the results of the item analyses, the data for the first ten participants (40% of the total number of participants) were coded and preliminary coding schemes were constructed, one coding scheme for each item type. The purpose of this step was to identify categories and

patterns of strategy use in relation to item types, and then use these categories and patterns in the form of coding schemes to code the rest of the data. Segments of a given verbal report were not coded as strategies unless they denoted strategic moves. As the coding process was proceeding, existing strategy categories were modified as necessary and new categories were added to the coding schemes.

The coding process was inductive, and so was data-driven in order to reveal strategies that are more specific and so more reflective of how participants in this study dealt with the reading tasks and items. This is to say that no strategies from previous research were imposed on data coding in this study. Throughout the coding process, the coding schemes were subjected to meticulous modifications in order to fine-tune the strategies that emerged during the initial coding and to fit emerging strategies into the schemes. Certain codes were rephrased so that they could describe strategies as precisely as possible. Other codes were combined when shown to share the same strategic theme or behavior with given item types.

During data collection and preparation for analysis, it was apparent that all participants dealt with the reading sets as if they had been part of an actual test-taking endeavor so any strategies they used were geared towards the provision of the right answer. In other words, participants can be said to have used far more test-taking strategies than reading strategies. Indeed, in all instances of strategy use where participants reported that they used certain reading strategies, such strategies were expressed as components of test-taking strategic moves. Codes were strategic moves that were either made explicit or implicit by the manner in which they or other strategies were verbalized. For example, one respondent reported his response to a BC-*v* item saying “I

could not understand the sentence, so I read the sentence before it and the one after it", suggesting that he read the sentence that had the vocabulary item first. Strategies were each assigned a tag or label for ease of reference in the quantitative and qualitative analyses. Each strategy tag or code identifies the task type to which the item belongs, the item type, and a serial number for the strategy to distinguish it from others used with the same item. For example, BC-*v*6 designates strategy number 6 as being among the strategies that participants used when answering the vocabulary items (*v*) that belong to the basic comprehension (BC) task type. (See Appendix M for abbreviations of task and item types used in strategy codes).

Owing to a high level of lucidity of strategic moves in the verbal data, assessment of intra-coder reliability was assumed to be more reasonable and accurate than inter-coder reliability. Another reason was that this research was conducted by a sole researcher, and so hiring an independent rater to estimate inter-coder reliability would have involved a rater who was either inexperienced with the coding of strategy data or unfamiliar with the nature of the data collected in this study, or both. Intra-coder reliability refers to the extent to which a single coder is consistent in coding qualitative data over the period of time or the number of times he has coded the data (A. Green, 1998).

In order to estimate intra-coder reliability in this study, a sample of episodes, constituting 40% of the total transcribed data, was randomly selected one month after all the verbal data were coded. The random selection made use of RAND function in Microsoft Excel. The reason for this step was to ensure that the selected episodes were representative of the transcribed data and that an estimate of coding reliability of the

sample data could be extrapolated to the rest of the data. The sample data that were recoded included a total of 160 episodes that had 460 segments. To reiterate, a segment represents a strategic move.

The statistic used to assess intra-coder reliability is Pearson's Product-Moment Correlation, which can be defined as a measure of the extent to which two variables are correlated or associated. Pearson *r* ranges from −1 to +1, with −1 indicating strongly negative relationship and +1 strongly positive relationship. The reason behind the choice of this statistic was its suitability for categorical data (Ark, Croon, & Sijtsma, 2005) and usefulness in assessment of inter-coder reliability (Bordens & Abbott, 2005). According to A. Green (1998), intra-coder reliability can be assessed using roughly the same measures as inter-coder reliability, including Pearson *r*. This calculation yielded an intra-coder reliability: $r\,(459) = 0.854$, $p<.01$, which points out a strongly positive, linear relationship between the first coding and the second coding of the sampled data. This, in turn, indicates a highly acceptable level of coding reliability. Codes that did not agree across the two codings were examined to ensure that each was accurately assigned to its respective segment; and as a result, these codes were either merged or discarded. The coding schemes were modified accordingly to reflect these decisions, with the aim being to augment coding reliability.

Procedures for Data Analysis

In data analysis, combining both quantitative figures and qualitative accounts of the data was deemed ideal to address the three research questions of this study. On the one hand, quantitative figures in terms of frequencies would indicate patterns of strategy use; on the other, qualitative accounts would describe in more detail how strategies relate

to such factors as item type, scoring level, and the choice of the right answer (cf., Creswell, 2009).

Quantitative analysis.

Data coding led to the development of a strategy list for each item type. Occurrences of strategies in the transcripts were tallied in order to calculate how frequently strategies were used with the given item types. The reason why occurrences of strategies were quantified was three-fold: First, to find out tendencies among participants with respect to what strategies they used frequently with which item types; second, to detect potential differences between top and low scorers regarding which group tended to use which strategies more frequently with the given item types than the other; and third, to explore aspects of effective strategy use through examination of clusters of frequent strategies that seemed to have resulted in the provision of the right answers to item types. Calculating frequencies of strategy use for each item type proved useful as each item called for the use of a set of strategies specific to its type.

Because the selected item types had among them pairs of BC-*v*, BC-*fi*, BC-*ss*, I-*bi*, I-*rp*, and I-*it* item types and single items representing the other item types, the former set of item types would presumably show higher strategy frequencies than the latter. To deal with this issue, raw totals of strategies were converted into type/token ratios by dividing the number of occurrences of the strategy by the number of items of a type selected for data analysis. For example, on both of BC-*v* items across the two reading sets, 18 participants reported that they read the sentence containing the target word and the surrounding context (or strategy BC-*v*4), so the type/token ratio for this strategy

equals the result of dividing 18 by 2, which is 9. Certain strategies were grouped together since they tended to occur together with a given item type.

Qualitative analysis.

This part of data analysis focused on offering qualitative accounts of strategy use by participants across item types, potential differences between high- and low-scoring participants in their strategy use, and aspects of strategy use that could be related to the choice of the right answer. That is, there were three goals of this analysis: First, to describe strategy use among participants as they were trying to answer each item type on the reading sets; second, to explain potential differences between high- and low-scoring participants in their strategy use; and third, to inductively identify and categorize themes that seemed to characterize effective strategy use. It was assumed that these qualitative accounts would reflect a clear and accurate picture of the nature of strategies that participants tended to use with the item types, potential discrepancies between the high and low scorers among participants in their strategy use, and aspects of effective strategy use that seemed to have helped test takers perform well in response to the task and item types.

At this stage, the data in both their current form and format seemed to speak to the three research questions in a straightforward fashion. For research question one, the kinds of strategies participants reported use of on the reading sets were each related to task and item types to determine what strategies were used with which task and item types. The scores the participants had on the reading sets were used to decide about those participants who could fit in the high- versus low-scoring groups to address the second research question. Strategy frequencies among the two scoring groups were used to

determine any potential differences between the two groups in strategy use. As for the third question, participants' successful answers to the question items on the reading sets were used to determine effective aspects of test-taking strategy use.

Summary

Chapter 3 described the research methodology used in this study, which involved the rationale for the research design and method, the selection of participants, materials, and procedures of data collection and analysis.

A combination of stimulated recall, self-observation, and retrospective interview procedures were used to collect the data of this study. This was the case in order to ensure the collection of data that are both all-inclusive and highly typical of what participants would do when they respond to the reading tasks of the actual TOEFL iBT. Appropriate measures in view of previous thought and research were taken to make the most of this procedural integration. Working with 25 participants in this study was expected to increase the chance of detecting potential patterns as well as discrepancies among test takers from the target population in regard to strategy use and performance on the TOEFL-iBT reading section. Each participant was asked to respond to two reading sets that were used as part of an actual TOEFL-iBT test administered previously by the ETS. Certainly, there would be concerns about the validity and reliability of the results of this study had the reading sets been selected from test preparation materials not published by the test makers.

Data collection procedures were carried out over three stages. First, the participant was oriented to the procedures and steps of data collection and trained in how to produce verbal reports. Second, the participant was engaged in the SRSORI procedure

in two sessions separated by a break. And third, the participant was debriefed about his or her performance and scoring, and rewarded for participating in the research. Before data collection, a pilot study was conducted in order to check research materials and procedures and apply to them any necessary adjustments on the basis of the outcomes of the pilot study. After all collected data were transcribed, based on the results of item analyses, verbal reports associated with certain items were selected for coding. A coding scheme was constructed for each item type by means of inductive coding of transcribed data. In data analysis, a combination of quantitative figures and qualitative accounts of the data was implemented in order to furnish inclusive and converging answers to the research questions.

Among the strengths of the methodological approach followed in this study are the following. The study employed a mixed-method procedure, which can be said to have augmented trustworthiness of the data collected (Flick, 2009). All procedures were used such that they called for minimal exertion on the respondents' part in order to keep them more focused on the tasks they were asked to perform. To ensure credibility, the researcher made every effort possible to ensure that participants would respond to the reading sets in almost the same manner they would on the actual TOEFL iBT (cf., Wallen & Fraenkel, 2006). To capitalize on both the quality and quantity of data collected in this study, the participants' interpretations and comments about their response behaviors were taken into consideration in data analysis and description, as is to be shown in Chapter IV.

It is the researcher's hope that the detailed description of the methodology followed in this study would enable readers to evaluate the procedures of data collection and analysis and future researchers to replicate this research (cf., Afflerbach, 2002).

Chapter IV presents the results and findings of this study according to research questions, in terms of test-taking strategy use across task and item types, test-taking strategy use among high- and low-scoring groups of test takers across task items, and aspects of effective test-taking strategy use across task and item types.

CHAPTER IV

FINDINGS

With the focus on how a sample of 25 Arab ESL learners behave in response to the reading section of the TOEFL iBT, this study attempted to answer three research questions: (a) What test-taking strategies do subjects use when responding to the reading tasks and items?, (b) Are there any differences between high- and low-scorers among subjects in their use of test-taking strategies on the reading tasks and items?, and (c) What aspects of effective test-taking strategy use do subjects tend to employ with the reading tasks and items? Strategy data used to address the three research questions were collected using a procedural integration of stimulated recall, self-observation, and retrospective interview.

This chapter presents the results and findings of the study in a manner arranged by the three research questions. The chapter is organized into four sections: (a) an overview of procedures for data analysis, (b) testing the research questions, (c) additional analysis, and (d) a summary of the chapter.

An Overview of Data Analysis

Frequencies of strategy occurrence in verbal reports were used to answer the three research questions: First, to find out tendencies among participants with respect to what strategies they used frequently with which item types; second, to detect potential differences between top and bottom scorers in regard to which group tended to use which strategies more frequently with item types than the other; and third, to explore aspects of effective strategy use through examination of clusters of frequent strategies that seemed to have resulted in the provision of the right answers to item types.

On the other hand, qualitative accounts of strategy use among participants are offered in three ways. First, strategies are explicated in relation to task and item types with which they were associated in verbal reports and exemplified with excerpts from the verbal data. Second, potential discrepancies between the high and low scorers in their strategy use are examined by focusing on the manner in which the two groups used strategies differentially in response to task and item types. And third, aspects of effective strategy use that seemed to have helped test takers answer item types correctly are described along with illustrations from verbal data.

Certain text styles and abbreviations are used in strategy descriptions and illustrations for the sake of clarity and brevity:

- *italic* text: Excerpts from verbal reports. The segment of an excerpt expressing the use of a given strategy is marked with <strategy#> (e.g., <*v*1>).
- Single quotation marks (‘’): Enclose an edited version of a participant’s own words in English.

- Double quotation marks (“ ”): Enclose content from the passages or the question items on the reading sets.
- Ellipsis mark (…): Used in excerpts to indicate the omission of a thought or expression unclear or irrelevant to the strategy instance being illustrated, a portion of the text unimportant to understanding the excerpt, or a researcher’s prompt or question.
- Abbreviations of task and item types as they appear in Appendix M.
- Strategy abbreviation (applicable throughout): Uses the abbreviation of the item type plus the strategy serial number (e.g., *v*3 denotes strategy 3 among strategies used with the vocabulary item type or BC-*v* items).

Testing the Research Questions

Research question one.

What test-taking strategies do subjects use when responding to the reading tasks and items? In order to address the first research question, an analytical approach similar to that of Cohen and Upton (2006) was followed, with the hope that this would ease comparison of findings across similar studies. Strategies that were used by participants across item types are sorted within their coding schemes in a descending order by their frequency of occurrence. Type/token ratios for strategies were used to categorize strategies into four levels of frequency rate in order to group strategies that were used with proximate rates of frequency. Ratios that resulted in values with decimals— all decimals that were 0.5 or higher—were rounded up to a whole integer. The reason for this step was to avoid reporting decimals, since such values did not show to have a significant effect on strategy distributions into the four levels of frequency rate. Taken

together, these steps helped discern notable trends and patterns of strategy use. The four levels of frequency rate are as follows:

- very high frequency (VH) ≥ 10
- high frequency (H) ≥ 5
- moderate frequency (M) ≥ 3
- low frequency (L) < 3

Strategies with frequency rates below 3 were not considered in the results and findings of the study. Besides their low values, such strategies did not show to reflect notable trends of strategy use or discrepancies between high- and low-scoring groups.

The presentation of the results for research question one appears in a format borrowed from Cohen and Upton (2006). Strategy trends and patterns are described in relation to item types, and illustrated with representative excerpts from verbal data.

Strategy abbreviation(s) specific to this section:

- (ID#RS#I#): Participant ID No., serial number of reading set (1 for Reading Set 1, and 2 for Reading Set 2), item number, respectively.

Strategy use by task and item type.

Basic comprehension-vocabulary (BC-v). BC-*v* is intended to "measure examinees' ability to comprehend the meanings of individual words and phrases as used in the context of the passage" (ETS, 2003, p. 4). A test taker needs to choose the option that represents the best synonym of the given word in its given context. (See Appendix C for an example of a BC-*v* item, as well as all the item types that are to be described). Table 6 below describes the most frequently used test-taking strategies with this item type.

Table 6

Common Strategies for Basic Comprehension: Vocabulary (BC-v)

Frequency Level	Strategy Code	Strategy Description	Frequency Rate
VH	v1	Read the sentence containing the target word (either to figure out the word meaning from the context or confirm the known meaning of the word).	15
	v4	Read the sentence containing the word and the surrounding context to figure out the word meaning.	10
H	v6	Eliminate certain options (either as incorrect, unknown, irrelevant, or unlikely).	9
	v5	Replace the word with each option to see which one best fits in the context.	6
	v2	Use clues to choose an option.	6
M	v9	Consult background knowledge (either to think of the word meaning or to choose an option).	4
	v7	Use guessing to choose an option.	4
	v8	Confirm the answer (either by trying the chosen option in place of the target word, checking other options, or rereading the context of the target word).	4
	v3	Read the options first (either to familiarize oneself with them or to choose a preliminary option).	4
	v10	Decide on an option (either by trying it in place of the target word, using a semantic clue, or guessing the meaning of the option).	4

A review of strategies in Table 6 reveals the following trends of strategy use for this item type:

1. Most test takers started response to BC-*v* items by reading the sentence containing the target word (*v*1), for instance, to confirm the meaning of the word that they knew from background knowledge: *I made a quick look at the word and the surrounding words to confirm my choice by verifying the meaning of the word in the context of geology* <*v*1> (ID12RS2I4). Some test takers felt they had to read the surrounding context besides the sentence containing the target word in order to figure out the meaning of the word (*v*4): *I read the sentence and the sentences before and after it* <*v*4> (ID4RS1I1). The surrounding context here refers to the sentence before the sentence containing the target word, or the sentence after it, or both.

2. There were three strategies that occurred with high frequency, including test takers' use of elimination of certain options (*v*6), for instance, when they found them to be incorrect in the light of the contextual meaning of the target word: *I eliminated Options 1 and 4 because the author is trying to praise President Andrew Jackson, but these two options do not mean praise* <*v*6> (ID18RS1I1). Another strategy that occurred at this level of frequency was the replacement of the target word with each option to determine which option best fitted in the context of the target word (*v*5): *I tried each one of the options* <*v*5>*, and tried Option 3 "remains"* <*v*5>*, but none fits in the context* (ID6RS2I4). The third strategy was the use of clues to choose an option (*v*2): *I paid attention to the period of time that extends from 1829 to 1837, it was not quite a short or long period. I connected this period* <*v*2>*to the meaning of the context. Therefore, I chose Option 3 "rapidly"* (ID16RS1I1).
3. The rest of the ten strategies included the use of background knowledge (*v*9), for instance, to think of the meaning of the target word: *I know 'measurably' from 'to measure'* <*v*9>*, and so "immeasurably" 'means something big'* (ID11RS1I1). Some test takers resorted to guessing to choose an option (*v*7): *I did not know Options 1and 4, but I guessed that the answer could be one of these two* <*v*7>*. So, I chose Option 1 because I know what Options 2 and 3 mean, but not Options 1 and 4* (ID25RS2I4).

   Also, some test takers confirmed their answers (*v*8), for instance, by replacing the target word with the answer to see how it sounded in context: *I found Option 3 "remains" to be the only one that gives the same meaning as "relics." I tried it and it fitted in the context* <*v*8> (ID22RS2I4). Test takers read the options first (*v*3), for instance, to familiarize themselves with them: *I read the options first to become*

familiar with them <v3>before I tried to find out the answer (ID9RS1I1). At moments of hesitation as to which option to choose from two or more options, test takers decided on an option (*v*10), for instance, by replacing the target word with each one of the options: *I was hesitant between Options 2 "greatly" and 3 "rapidly," but decided on Option 3 later based on the context of the sentence. I tried replacing the* [target] *word with this option and it sounded OK to me* <*v*10> (ID21RS1I1).

Basic comprehension-factual information (BC-fi). BC-*fi* is intended to "measure examinees' ability to identify responses to questions about factual information that is explicitly stated in a text" (ETS, 2003, p. 10). A test taker needs to choose the option that has the information content shared by the option and the text and sought by the item stem. (See Appendix C for an example of a BC-*fi* item). Table 7 below describes the most frequently used test-taking strategies with this item type.

Table 7

Common Strategies for Basic Comprehension: Factual Information (BC-fi)

Frequency Level	Strategy Code	Strategy Description	Frequency Rate
VH	fi1	Read the paragraph or the relevant portion of it in order to locate critical information.	19
	fi6	Eliminate certain options (either as incorrect, irrelevant, incomplete, or unmentioned).	10
H	fi5	Read the options first (either to familiarize oneself with the options or to use them to locate critical information in the paragraph).	6
	fi2	Confirm the answer (either by rereading the paragraph or the relevant portion of it, or checking other options).	5
M	fi7	Match key word(s) or phrase(s) in the item stem or options and the paragraph.	4
	fi8	Consult background knowledge (either to consider an option, to choose a preliminary option, or to understand the item content).	4
	fi4	Use clues or guessing to choose an option.	4
	fi3	Reread the paragraph or a portion of it (either to locate the critical information or to decide on an option).	4

A review of strategies in Table 7 reveals the following trends of strategy use for this item type:

1. Test takers made frequent use of strategy (*fi*1) which involved reading the paragraph referred to by the item or the relevant portion of it:

   > *I read the paragraph < fi*1>. *I used understanding of the paragraph and background knowledge in geography like when plates collide, mounts are formed. In the paragraph, it says "plates crashing into each other," and Option 3 is almost the same idea* (ID1RS2I5).

   Some test takers chose to read a portion of the paragraph that is relevant to the intent of the question, because they had already read it while answering the previous item: *I skimmed the paragraph fast starting from the point where I stopped when I read it for the previous item, which was "Government's role …" <fi*1>*, trying to understand* (ID16RS1I7). Test takers also used the strategy of eliminating certain options (*fi*6), for instance, when they found these options to be irrelevant to the intent of the question: *When I started, I read the paragraph to find the answer. I eliminated Option 1 because it has no connection* [with the intent of the question] <*fi*6> (ID7RS1I7).

2. Two strategies were used at high frequency among test takers. First, test takers started their response to the BC-*fi* items by reading the options (*fi*5), for instance, to use them to locate the source of critical information in the paragraph: *Since the question asks "… which of the following areas?", I skimmed the options and read the paragraph while using the options to find the answer <fi*5> (ID6RS1I7). The other strategy at this level of frequency was confirmation of the chosen answer (*fi*2), for instance, by rereading the paragraph: *I read the options first and chose Option 3. Then, I reread the paragraph to confirm my answer <fi*2> (ID12RS1I7).

3. There were four strategies that occurred at a moderate level of frequency. The first strategy involved matching key words or phrases in the item stem or options, and the paragraph (*fi*7): *I matched a key word "religion" in the paragraph and an option <fi*7*>… I chose Option 1 because the word "religion" appears in the paragraph and the option* (ID16RS1I7). The second strategy in this category was (*fi*8) by means of which test takers used background knowledge, for instance, to choose a preliminary option: *I have background about this topic, geography and the like, and so I expected Option 2 to be the right answer <fi*8*>* (ID24RS2I5).

Strategy three at the moderate level of frequency was strategy (*fi*4) which involved the use of clues or guessing to choose an option, for instance, by using the meaning of a key word:

> *I read in the paragraph "Whigs and Democrats differed not only in their attitudes toward the market but also about how active …." This combines the two ideas that suggest "moral beliefs"* [Option 4]*, so I chose Option 4 using "attitudes" as 'a key word' <fi*4*>*. (ID5RS2I5)

The fourth strategy at this level was rereading the paragraph referred to by the item or a portion of it (*fi*3), for instance, to decide on an option in case of hesitation: *I know the topic. I know that 'tectonic plates move,' but I had to reread the paragraph because I got dizzy with the way the options are phrased <fi*3*>* (ID10RS2I5).

Basic comprehension-negative fact (BC-nf). BC-*nf* is intended to "measure examinees' ability to verify what information is true and what information is NOT true or not included in the passage based on information that is explicitly stated in the passage" (ETS, 2003, p. 12). A test taker needs to choose the option that is not supported by the information given in the paragraph referred to by the BC-*nf* item. (See Appendix C for an example of a BC-*nf* item). Test takers responded to only one BC-*nf* item which is Item 10

on Reading Set 1. Table 8 below describes the most frequently used test-taking strategies with this item type.

Table 8
Common Strategies for Basic Comprehension: Negative Fact (BC-nf)

Frequency Level	Strategy Code	Strategy Description	Frequency Rate
VH	nf1	Attempt to locate each option in the paragraph while eliminating certain options.	10
	nf4	Read or skim the paragraph while eliminating certain options (either as incorrect, irrelevant, or mentioned in the paragraph).	10
H	nf5	Read the options first (either to familiarize oneself with them or to choose a preliminary option).	9
	nf2	Read the paragraph to figure out the exception.	8
M	nf3	Confirm the answer (either by rereading the paragraph or a portion of it, or ensuring elimination of the wrong options).	4

A review of strategies in Table 8 reveals the following trends of strategy use for this item type:

1. The most frequently used strategies with the BC-*nf* item were considerably specific to this item type. Strategy (*nf*1) was aimed to locate each option in the paragraph and eliminate it if it exists. It called for the test taker's switching between the options and the paragraph: *I read the options first … I switched between the paragraph and the options, trying to determine if an option is mentioned in the paragraph or not* <*nf*1> (ID6RS1I10). Test takers who used strategy (*nf*4) read or skimmed the paragraph while eliminating certain options, for instance, when such options sounded irrelevant to the intent of the question:

> *The paragraph talks about two categories of people and the question asks about the first category or the Democrats, so each option of these four can either be about the Democrats or the public. I eliminated those options that talk about the public like Option 1 <nf4>. (ID21RS1I10)*

2. At the high level of frequency, test takers used two strategies with BC-*nf*. The first strategy involved reading the options first (*nf*5), for instance, in order to have familiarity with options that could guide their determination of the exception when they read the paragraph. Thus, this strategy was often followed by the strategy of reading the paragraph: *I read the options <nf*5*>. I read the paragraph, trying to locate information on Democrats and focus on this information, because that is what the question is asking about* (ID21RS1I10). The second strategy was reading the paragraph to determine the exception among the options (*nf*2): *I preferred to read the whole paragraph because it* [the item] *is a negative fact. I focused on the given groups and whether a given group supports or opposes <nf*2*>* (ID4RS1I10).
3. Only one strategy was used at the moderate level of frequency. In their use of strategy (*nf*3), test takers confirmed their answers to the BC-*nf* item, for instance, by checking the options to make sure that all wrong options were eliminated: *Option 3 "rising entrepreneurs" is the only option that is not clearly stated in the paragraph. I reread the options to ensure that I cancelled out the wrong options <nf*3*>* (ID17RS1I10).

Basic comprehension-pronoun reference (BC-pr). BC-*pr* is intended to "measure examinees' ability to identify relationships between pronouns and other anaphoric devices and their antecedents/postcedents within the passage" (ETS, 2003, p. 6). A test taker needs to choose the option that represents the referent of the pronoun. (See Appendix C for an example of a BC-*pr* item). Test takers responded to only one BC-*pr* item which is Item 8 on Reading Set 2. Table 9 below describes the most frequently used test-taking strategies with this item type.

Table 9

Common Strategies for Basic Comprehension: Pronoun Reference (BC-pr)

Frequency Level	Strategy Code	Strategy Description	Frequency Rate
VH	pr1	Read the sentence containing the target pronoun to identify the referent.	12
H	pr2	Confirm the answer (either by rereading the sentence containing the pronoun or the relevant portion of the paragraph, or checking other options).	8
M	pr4	Read the sentence containing the pronoun more than once to identify the referent.	4
	pr5	Use clues to choose an option.	3
	pr3	Eliminate certain options as incorrect.	3

A review of strategies in Table 9 reveals the following trends of strategy use for this item type:

1. Test takers used only one strategy at the very high level of frequency. Using strategy (*pr*1), test takers read the sentence that contains the pronoun in order to identify the referent of the pronoun: *I read the line: "…, and these slowly moving masses of ice cut out valleys, carrying with them" <pr*1*>*; *"with them" means "masses of ice." I understood "them" to mean "masses of ice" or Option 2* (ID11RS2I8).
2. There was also a sole strategy at the high level of frequency. Test takers had a noticeable preference to confirm their answers to this item (*pr*2), for instance, by rereading the relevant portion of the paragraph:

   > *I read the sentence that has the pronoun, and the answer was clear from "and these slowly moving masses of ice cut out valleys, carrying with them." This item is more of 'a grammar item.' I chose Option 2 "masses of ice." I continued reading and reread the surrounding context to confirm my answer <pr*2*>*. (ID24RS2I8)

3. Test takers used three strategies at the moderate level of frequency. The first strategy involved reading the sentence that contains the target pronoun more than one time in order to identify the referent (*pr*4): *I read the paragraph from the beginning and noticed that the author starts talking about "glaciers." I reread starting from the*

sentence that has the pronoun <pr4>, and found "glaciers" to be the referent (ID10RS2I8). When using the second strategy at this level, test takers used clues to choose an option (*pr*5): *I do not know "glaciers." I used the comma before "carrying with them" <pr*5*>, which indicates that "them" refers to the last word before the comma* [i.e., "valleys" or Option 3] (ID8RS2I8). In another instance of strategy use at this level, test takers eliminated the options that they found to be incorrect based on their reading of the sentence that has the target pronoun (*pr*3): *I read the whole sentence to understand the full context and pronoun references. Through my first reading, I felt that the answer must be one of the first three options. I eliminated Option 1 "cold areas"… because you need to base your choice on a previous word, and the last word* [phrase] *"rock debris" <pr*3> (ID4RS2I8).

Basic comprehension-sentence simplification (BC-ss). BC-*ss* is intended to "measure examinees' ability to identify essential information as they process complex sentences in extended texts without getting lost in less important details and elaborations" (ETS, 2003, p. 8). A test taker needs to choose the option that represents a simplified form of the target, highlighted sentence, which maintains the central idea of the sentence. (See Appendix C for an example of a BC-*ss* item). Table 10 below describes the most frequently used test-taking strategies with this item type.

A review of strategies in Table 10 reveals the following trends of strategy use for this item type:

1. Only one strategy was used with BC-*ss* at the very high level of frequency. Strategy (*ss*1) was one of the most frequent strategies across all item types. Using this strategy, test takers read the highlighted sentence (*ss*1): *I read the highlighted sentence <ss*1>

twice trying to comprehend or grasp as many pieces of information as I could (ID2RS1I11).

Table 10

Common Strategies for Basic Comprehension: Sentence Simplification (BC-ss)

Frequency Level	Strategy Code	Strategy Description	Frequency Rate
VH	ss1	Read the highlighted sentence.	25
H	ss4	Eliminate certain options (either as incorrect, irrelevant, ambiguous, or narrow).	9
	ss3	Match the content of the highlighted sentence and the options.	5
	ss8	Read the highlighted sentence while focusing on its main points.	5
	ss2	Use clues to choose an option.	5
M	ss5	Decide on an option using a semantic clue.	4
	ss9	Confirm the answer (either by rereading the highlighted sentence or checking other options).	4
	ss7	Use understanding of the general meaning of the highlighted sentence to choose an option.	4
	ss6	Read the highlighted sentence more than once to further understanding.	3
	ss10	Use guessing to choose an option.	3

2. Test takers used four strategies that can be considered highly frequent. The most frequent one of the four strategies was elimination of options (*ss*4), for instance, when such options sounded incorrect in the light of the information content of the highlighted sentence: *I eliminated Option 2 because 'the areas here were not naturally split, because if the Democrats had been able to go to the cities and towns, they would have gone. I do not think that this option is right'* <*ss*4> (ID7RS1I11). The second among the highly frequent strategies was matching the content of the highlighted sentence and the options to identify what would potentially be the right option (*ss*3): *I read the highlighted sentence, and matched words in the highlighted sentence and the options … I preferred Option 3 because it has "more isolated" similar to "increasingly isolated" in the highlighted sentence* (ID11RS1I11).

At the same rate of frequency as strategy (*ss*3), was strategy (*ss*8) which involved test takers' reading of the highlighted sentence with the intention of focusing on its main points: *I read the highlighted sentence. I focused on two main ideas: 'how they resist forces of destruction and how they have a relatively short life'* <*ss*8> (ID24RS2I9). The fourth strategy at this level was the use of clues to choose an option (*ss*2): *I read the options, and chose Option 2 using "although" as a key word which implies a sense of contradiction as does "but" in the paragraph* <*ss*2> (ID25RS2I9).

3. At the moderate level of frequency, five strategies were used by test takers. The most frequent strategy at this level was deciding on an option (*ss*5) by means of semantic clues: *I hesitated between Options 2 and 3 because both suggest a short period of time. I favored Option 3 which has "but only for a short time," because in the paragraph, it says "relatively short-lived in geological terms"* <*ss*5> (ID18RS2I9). The second most frequent strategy at this level was confirming the answer (*ss*9), for instance, by checking the other options to ensure that none of them was better than the chosen one: *Option 2 is very close to the highlighted sentence, and because of this, I chose it directly. I read Options 3 and 4 but they are not as close to the answer as Option 2* <*ss*9> (ID12RS2I9).

   Strategy (ss7) comes third at this level, which called for test takers' use of their understanding of the general meaning of the highlighted sentence to choose an option: *I chose Option 4 because it was the closest of all options based on the general meaning of the highlighted sentence* <*ss*7> (ID19RS2I9). Some test takers felt they had to read the highlighted sentence more than one time to further their understanding

of it (*ss*6): *I read the highlighted sentence more than once* <*ss*6> *because I did not read the passage or the paragraph 'to understand everything'* (ID10RS2I9). Strategy five at this level involved the use of guessing to choose an option (*ss*10): *No much time was left, so I chose* [an option] *quickly and randomly* <*ss*10> (ID8RS1I11).

Inferencing-basic inference (I-bi). I-*bi* is intended to "measure examinees' ability to comprehend an argument or an idea that is strongly implied but not explicitly stated in the text" (ETS, 2003, p. 25). A test taker needs to choose the option that best reflects what can be inferred from given information in the paragraph referred to by the I-*bi* item. (See Appendix C for an example of an I-*bi* item). Table 11 below describes the most frequently used test-taking strategies with this item type.

Table 11
Common Strategies for Inferencing: Basic Inference (I-bi)

Frequency Level	Strategy Code	Strategy Description	Frequency Rate
VH	bi2	Read the paragraph to locate critical information.	23
	bi4	Eliminate certain options (either as incorrect, irrelevant, unclear, or nonmeaningful).	11
H	bi1	Read the options first (either to familiarize oneself with them, to choose a preliminary option, or to locate critical information in the paragraph).	6
	bi3	Reread the paragraph or a portion of it (either to further understanding, to check an option, or to decide on an option).	5
	bi6	Attempt to locate the critical information using a key word/phrase.	5
	bi5	Use clues or guessing to choose an option.	5
	bi8	Confirm the answer (either by rereading the paragraph or the relevant portion of it, or checking other options).	5
M	bi7	Match the content of the paragraph and the options.	4

A review of strategies in Table 11 reveals the following trends of strategy use for this item type:

1. Test takers used two strategies at the very high level of frequency. First, they read the paragraph referred to by the I-*bi* item while trying to locate the critical information (*bi*2) that they could use to make the required inference: *I read the paragraph … I preferred Option 1 because of the second sentence in the paragraph* <*bi*2> (ID2RS1I9). Second, test takers used elimination of options (*bi*4), for instance, when they found such options incorrect based on what they understood from reading the paragraph: *I skimmed the paragraph … I eliminated some options … Option 3 is not correct because it changes the number mentioned in the paragraph to "400 million years ago" when it is supposed to be "50 million years"* <*bi*4> (ID21RS2I3).
2. Out of the eight common strategies used with the I-*bi* items, five strategies were used at the high level of frequency. There were test takers who preferred to read the options first (*bi*1), for instance, in order to use them to locate the critical information in the paragraph: *I read the options to know what I have to find in the paragraph and where I can find it* <*bi*1> (ID21RS2I3). Second, there was a noticeable tendency among test takers to reread the paragraph referred to by the I-*bi* item or a portion of it (*bi*3), for instance, in order to further understanding of the information presented in the paragraph: *I read the paragraph more than once, trying to understand it* <*bi*3> (ID10RS1I9).

   When using strategy number three at this level, test takers attempted to locate the critical information in the paragraph using a key word or phrase in an option (*bi*6): *I read the options first. Then, I scanned the paragraph for the word "Himalayas"* <*bi*6> (ID6RS2I3). Test takers made their choice of options using clues or guessing (*bi*5): *The last sentence has "favor social-reform legislation and aid to*

education," and that is why it encouraged me to choose Option 4 because it has "interests" meaning attraction or benefit <*bi*5> (ID1RS1I9). Strategy number five at this level involved confirmation of the answer (*bi*8), for instance, by rereading the relevant portion of the paragraph:

> *I felt I guessed the answer, so I scrolled up to make sure I got the right answer. I skimmed this part of the paragraph and found it to be about "public liberty"* <*bi*8>*, so I confirmed my answer and felt comfortable.* (ID17RS1I9)

3. At the moderate level of frequency, test takers used only one strategy. Using strategy (*bi*7), test takers matched the content of the paragraph and the options so as to identify the option that would possibly be the right answer:

> *I read the paragraph, trying to find the options. I found "They believed that it should be used to protect individual rights and public liberty" in the paragraph … and found Option 1 saying "They were focused on issues of public liberty"* <*bi*7>*. Therefore, this* [Option 4] *was my answer to the question.* (ID18RS1I9)

Inferencing-insert text (I-it). I-*it* is intended to "measure examinees' ability to understand the lexical, grammatical, and logical links between successive sentences" (ETS, 2003, p. 31). A test taker needs to insert a sentence into one slot out of four slots marked with (■), where the sentence would best fit—each slot representing an option for the I-*it* item. (See Appendix C for an example of an I-*it* item). Table 12 below describes the most frequently used test-taking strategies with this item type.

A review of strategies in Table 12 reveals the following trends of strategy use for this item type:

1. Test takers used only one strategy at the very high level of frequency. All test takers started response to the I-*it* items by reading the sentence which they should insert into a slot in the provided paragraph (*it*1), henceforth, the given sentence: *I tried to read the whole sentence* <*it*1> (ID1RS2I11).

Table 12

Common Strategies for Inferencing: Insert Text (I-it)

Frequency Level	Strategy Code	Strategy Description	Frequency Rate
VH	it1	Read the given sentence.	25
H	it2	Read the paragraph to determine where the given sentence best fits.	9
	it3	Use a logical connection between the given sentence and the sentence(s) before and/or after a given slot to choose an option.	8
	it4	Try the given sentence in one or more of the slots in the paragraph to determine where it best fits.	7
	it5	Eliminate certain options (if the surrounding context is either distant from the context of the given sentence, irrelevant to the theme of the given sentence, or composed of sentences that are logically connected).	6
	it8	Use clues or guessing to choose an option.	6
	it6	Confirm the answer (either by rereading the relevant portion of the paragraph, or checking other options).	5
M	it7	Decide on an option (either by rereading the paragraph or a portion of it, using key words or phrases, or using clues).	4

2. At the high level of frequency, six out of the eight strategies in Table 12 were used by test takers. First, test takers using strategy (*it*2) read the paragraph referred to by the I-*it* item in order to determine where the given sentence best fitted: *I read the paragraph while thinking where I could insert the sentence* <*it*2> (ID4RS1I12). Second, test takers tried to figure out a logical connection between the given sentence and the sentence(s) before and/or after a given slot in order to choose an option (*it*3): *I found the best slot to be Option 1 because* [the first sentence in the paragraph says] *"during Jackson's second term, the Whig party was formed" and the given sentence states "This new party." It defines the Whig party* <*it*3> (ID6RS1I12). There were test takers who preferred to try the given sentence in one or more of the slots in the paragraph to see where it would best fit (*it*4): *I tried the given sentence in all of the four places* [slots] <*it*4> (ID17RS2I11).

Test takers eliminated certain options (*it*5) on I-*it*, for instance, if the context of the slot comprises sentences that are logically connected without the given sentence: *I eliminated Option 2 because it says "It carries …," which means that the sentence here refers to something that is mentioned before* <*it*5> (ID23RS2I11). Another strategy used at the high level of frequency was the choice of an option by means of clues or guessing (*it*8):

> *I read the paragraph while focusing on the places* [slots]. *I got puzzled with it* [this item], *as I was paying attention to the time. Therefore, I was making more of scanning than reading. At the end, I guessed the answer* <*it*8>. (ID24RS1I12)

There were cases among test takers who tended to use a strategy of confirming their answers on the I-*it* items (*it*6), for instance, by checking the other options because of the possibility that one of them would prove better than the selected option: *I tried the other options to see if the* [given] *sentence could fit in another place* <*it*6>. *I saw that the given sentence fits only in slot 4* [Option 4] (ID16RS2I11).

3. Only one strategy was used at the moderate level of frequency. Upon hesitation as to which option to choose, test takers had to decide on an option (*it*7), for instance, by rereading a portion of the paragraph: *I hesitated between Options 2 and 3. Then, I tried to reread these sentences* [pointing to the bottom portion of the paragraph] *to decide on one of these two options* <*bi*7> (ID18RS2I11).

Inferencing-rhetorical purpose (I-rp). I-*rp* is intended to "measure examinees' ability to identify the author's underlying rhetorical purpose in employing particular expository features in the passage and in ordering the exposition in a particular way. Correct responses require proficiency at inferring the nature of the link between specific features of exposition and the author's rhetorical purpose" (ETS, 2003, p. 27). A test

taker needs to choose the option that best expresses the author's purpose of discussing an idea or using an example. This idea or example is marked in the text with a highlight. (See Appendix C for an example of an I-*rp* item). Table 13 below describes the most frequently used test-taking strategies with this item type.

Table 13
Common Strategies for Inferencing: Rhetorical Purpose (I-rp)

Frequency Levels	Strategy Code	Strategy Description	Frequency Rate
VH	rp6	Read the paragraph to figure out required information.	14
	rp2	Read the sentence containing the highlighted phrase, and the surrounding context if needed, to figure out required information.	13
H	rp5	Eliminate certain options (either as incorrect, irrelevant, ambiguous, or distant from the context of the highlighted phrase).	8
	rp1	Read the options first (either to familiarize oneself with them, to use them to locate critical information, to figure out required information, or to choose a preliminary option).	7
	rp3	Read the paragraph or a portion of it more than once.	5
M	rp8	Match the content of the sentence containing the highlighted phrase and the options.	4
	rp9	Decide on an option by rereading the paragraph.	3
	rp4	Confirm the answer (either by rereading the paragraph or the relevant portion of it, or checking other options).	3
	rp7	Use clues or guessing to choose an option.	3

A review of strategies in Table 13 reveals the following trends of strategy use for this item type:

1. Two strategies were used at the very high level of frequency with the I-*rp* items. The most frequent strategy in this category was reading the paragraph to figure out the required information (*rp*6): *I read the paragraph and figured out that "bankers and investors" is an example on a previous sentence in the paragraph that talks about how some people controlled the economy of the country* <*rp*6> (ID23RS1I3). The second most frequent strategy was reading the sentence containing the highlighted

phrase, and the surrounding context if needed, to figure out the required information (*rp*2): *I read from the beginning of the sentence <rp2>, but it was not enough, so I read one sentence back, trying to understand the subject and deduce what the author means by the highlighted phrase <rp2>* (ID6RS1I3).

2. There were three strategies that occurred with high frequency, including test takers' use of elimination of certain options (*rp*5), for instance, when they found these options to be distant from the context of the highlighted phrase: *All options are available in the paragraph, so I used how each one is distant from "Carbon dioxide"* [the highlighted phrase] *to decide which options to eliminate <rp5> and which ones to keep* (ID9RS2I6). Another strategy at this level of frequency was reading the options first (*rp*1), for instance, in order to use them to locate the critical information in the paragraph: *I read the options <rp1>… I read the sentence that has the highlighted phrase while trying to find information that matches one of the options* (ID24RS1I3). Some test takers had to read the paragraph, or a portion of it more than once (*rp*3): *I read this part of the paragraph* [the sentence containing the highlighted phrase and the surrounding context] *once and twice <rp3>* (ID21RS2I6).
3. At the moderate level of frequency, four out of the nine strategies in Table 13 were used by test takers. These strategies include matching the content of the sentence with the highlighted phrase and the options (*rp*8):

> *I tried to find information that links between the paragraph and one of the options. I found that "rich" is the key word that has to do with money and banking, and it is the word that links between "bankers and investors"* [the highlighted phrase] *and "becoming rich"* [in Option 2] <*rp*8>. (ID18RS1I3)

At moments of hesitation as to which option to choose from two or more options, test takers decided on an option (*rp*9), for instance, by rereading the paragraph: *I hesitated*

*between Options 1 and 3. After reading the paragraph one more time, I chose Option 1 <rp*9> (ID17RS2I6).

There were test takers who confirmed their answers (*rp*4), for instance, by checking the other options:

> *I understood the meaning. "carbon dioxide" is the result of a process, which makes it clear that the author tries "to explain the origin of a chemical that can erode rocks"* [Option 1]. *I checked all of the other options <rp*4*>, but none seems to fit in the context except Option 1.* (ID24RS2I6)

Some test takers resorted to the use of clues or guessing to choose an option (*rp*7): *I tried to link between information in the paragraph and the options, but did not understand because of unknown vocabs in the options, for example, "erode rocks"* [Option 1], *so I chose an option randomly <rp*7> (ID6RS2I6).

Reading to learn-prose summary (R2L-ps). R2L-*ps* is intended to "measure examinees' ability to understand the major ideas and relative importance of information in a text … An introductory sentence is provided, and examinees select 3 additional sentences from 6 options … [The three correct options] represent the major ideas in the text that, taken together, form a high-level summary of the text" (ETS, 2003, p. 15). A test taker needs to select three answer choices that best summarize the passage, which demands that a test taker develop understanding by reading through the whole passage. (See Appendix C for an example of a R2L-*ps* item). Table 14 below describes the most frequently used test-taking strategies with this item type.

A review of strategies in Table 14 reveals the following trends of strategy use for this item type:

1. Three strategies were used with the R2L-*ps* item at the very high level of frequency. The first strategy was the selection of answer choices that test takers considered as

important points of the passage or points that offered a summary of the passage (*ps*6): *I read the answer choices. I used understanding from working through the previous items and skimming of the passage to select the best three answer choices that are the most important* [points] *in the passage* <*ps*6> (ID6RS1I13). Some test takers used the understanding they had developed as a result of going through the previous items and reading their respective paragraphs to select answer choices (*ps*2): *I read the answer choices while trying to remember if any of the ideas I understood from answering previous items* <*ps*2> are *repeated here* (ID1RS1I13). There were test takers who employed an elimination strategy (*ps*3) with certain answer choices, for instance, when such answer choices sounded irrelevant to the subject matter of the passage:

> *I read the answer choices one by one … I eliminated Choice 2* ["During Andrew Jackson's two terms as President …"], *because the passage does not mainly talk about this, but rather how the two parties dealt with people and economy, and how they collected wealth* <*ps*3>. (ID7RS1I13)

Table 14

Common Strategies for Reading to Learn: Prose Summary (R2L-ps)

Frequency Levels	Strategy Code	Strategy Description	Frequency Rate
VH	ps6	Select answer choices that represent important or summary points of the passage.	13
	ps2	Use understanding from going through the previous items and reading their respective paragraphs.	13
	ps3	Eliminate certain answer choices (either as incorrect, irrelevant, incomplete, or inappropriate to be summary points).	11
H	ps4	Use random guessing to select answer choices.	7
	ps5	Use the introductory sentence to determine required information.	6
	ps1	Use understanding from skimming or reading of the passage.	5
M	ps7	Confirm selection of answer choices (either by rereading the selected answer choices or checking the unselected answer choices).	4
	ps8	Use rational guessing to select answer choices.	3

2. Three strategies occurred at the high level of frequency for this item type. First, test takers made their selection of answer choices by means of random guessing (*ps*4): *I suspected that both Options 1 and 2 are correct, so I chose Option 2, taking a risk because I was not sure* <*ps*4> (ID11RS1I13). The second strategy in this category involved the use of the introductory sentence to determine which answer choices would represent the required information (*ps*5): *I read the given* [*introductory*] *sentence to see what the author wants in general* <*ps*5> (ID2RS1I13). Strategy number three at this level called for test takers' use of the understanding they had developed as a result of their reading or skimming of the passage (*ps*1):

   > *I thought I had to read the passage in such a way that is open* [i.e., unrestrained by care to understand specific details] *or general enough to help answer this item because this item asks about the main points* <*ps*1>. *So, I skimmed the passage in order to keep the main points in the back of my mind till I get to this item* <*ps*1>. (ID17RS1I13)

3. The two remaining strategies with the I-*ps* item were used at a moderate level of frequency. The first strategy involved test takers' confirmation of their selections (*ps*7), for instance, by rereading the selected answer choices: *I noticed the time warning, so I decided to select any answer choices at random. But then, I felt I could read the answer choices to check my selections* <*ps*7>, *and they appeared OK to me* (ID10RS1I13). Test takers using the second strategy at this level selected answer choices using rational guessing (*ps*8): *I felt comfortable about Choice 5 'the fundamental difference' because it is oft-repeated in the passage, especially in the last paragraph* <*ps8*> (ID24RS1I13).

Reading to learn-schematic table (R2L-st). R2L-*st* is intended to "measure examinees' ability to conceptualize and organize major ideas and other important

information from across the text … Correctly completed formats of these types reflect an able reader's mental framework of the text" (ETS, 2003, p. 18). A test taker needs to select answer choices from a list of statements, and fit these choices in a table of two distinct, major themes discussed in the passage. The R2L-*st* demands that a test taker be acquainted with defining elements that typify each theme. (See Appendix C for an example of a R2L-*st* item). Table 15 below describes the most frequently used test-taking strategies with this item type.

Table 15

Common Strategies for Reading to Learn: Schematic Table (R2L-st)

Frequency Levels	Strategy Code	Strategy Description	Frequency Rate
VH	st3	Use understanding from going through the previous items and reading their respective paragraphs.	10
H	st2	Refer to the passage to determine which processes are "constructive" and which are "destructive".	9
	st4	Confirm the selected answer choices (either by referring to the passage or checking the unselected answer choices).	6
	st6	Eliminate certain answer choices (either as incorrect, irrelevant, or unmentioned in the passage).	6
	st1	Use random guessing to select answer choices.	6
M	st7	Use understanding from skimming or reading of the passage.	4
	st5	Use judgment to decide whether a single process is "constructive" or "destructive".	4
	st8	Select answer choices that are certain or direct, or share certain thematic feature(s).	4

A review of strategies in Table 15 reveals the following trends of strategy use for this item type:

1. Only one strategy was used with the R2L-*st* item at the very high level of frequency. Test takers who used this strategy used the understanding they had developed as a result of going through the previous items and reading their respective paragraphs to

select answer choices (*st*3): *Through my answers to the previous items <st*3*>, I knew that Choice 5 is a destructive process* (ID6RS2I12).

2. Four strategies were used at the high level of frequency for this item type. First, test takers preferred to consult the passage to check certain answer choices when they felt uncertain about whether a given answer choice represents a "constructive" or "destructive" process (*st*2): *I read all the answer choices, trying to keep them in my mind. I referred to the passage to make sure as to which choices are "constructive" and which are "destructive" <st*2*>* (ID24RS2I12). Second, test takers who used strategy (*st*4) confirmed the selections they had made, for instance, by referring to the passage: *From my answers to the first or the second item about how mounts are formed, I mean "constructive processes" … The 2nd constructive process here is Choice 6. I went to the passage to confirm Choice 6 <st*4*>* (ID1RS2I12).

   Strategy number three at this level of frequency was elimination of answer choices (*st*6), for instance, when such answer choices sounded incorrect based on what test takers understood: *Choice 2 is mentioned when the author discusses the "Caledonian mountains," but the author does not treat it as one of the factors <st6>* (ID11RS2I12). Test takers using strategy (*st*1) resorted to random guessing to select an answer choice: *Under the pressure of time, I had to pick anything <st*1*>… I made a quick selection of answer choices* (ID7RS2I12).

3. At the moderate level of frequency, three strategies were used with R2L-*st*. The first of these strategies was the use of the understanding that test takers had developed from skimming or reading the passage to select answer choices (*st*7): *I used my knowledge from reading the passage <st*7*>… to determine which processes are*

"constructive" and which are "destructive" (ID17RS2I12). When test takers felt hesitant about whether a single process was "constructive" or "destructive", they used their judgment by imagining how a given process would affect Earth's landscape (*st*5):

> *I tried to find examples of these processes* [in the passage], *but did not find any examples and there is a lot of information, so I thought I should rely on myself <st*5*>. I figured out that "uplifts" are like Choices 1, 2, 5, and 6, and for the weather conditions, I selected 7 and 3.* (ID10RS2I12)

Test takers used strategy (*st*8) at this level to select answer choices, using the extent to which answer choices are certain or direct:

> *I remembered that there are three causes of mountain formation: The movement of the Earth's crustal plates… "Earthquakes"* [Choice 5] *is mentioned in a direct way <st*8*>; it is listed as one of the reasons for mountain formation … "Wind-driven sand"* [Choice 3] *is sure <st*8*>*. (ID23RS2I12)

Test takers also selected answer choices that seem to share a certain thematic feature (*st*8):

> *I did not know the meaning of "constructive," so I selected the three choices that are linked together; they are all phenomena that are naturally related <st*8*>, I mean Choice 1 "Collision of Earth's crustal plates," Choice 6 "Volcanic activity," and Choice 5 "Earthquakes."* (ID25RS2I12)

Discussion.

Research question one (*What test-taking strategies do subjects use when responding to the reading tasks and items?*) explores test-taking strategies that participants in this study tend to use when responding to the reading tasks and items on the TOEFL-iBT reading section. The answer to the first research question takes the form of strategy trends and patterns as indicated through the use of three levels of frequency.

The test-taking strategies used by test takers with each item type were fitted into three groups. The first group of strategies represented the most frequently used strategies.

The use of these strategies was generally not unexpected considering the formats of the item types. The formats of the test items were different across the item types, which called for test takers to use strategies that conformed to the demands of given item types. This finding is consistent with those of previous studies (e.g., N. Anderson et al., 1991; Rupp et al., 2006; Storey, 1997; Tsagari, 1994) which indicated that formats of test items determined the kinds of strategies test takers tended to use in these studies. The use of strategy (*bi*2) with the I-*bi* items at a rate of frequency of 23 attests to this fact. Almost 92% of test takers started their response to the I-*bi* items by reading the paragraph referred to by the item in order to locate the critical information that could help them figure out the answer. Consideration of the I-*bi* item format would show that this item type cannot be answered correctly without reading the paragraph the item refers to.

What the most frequent strategies also seem to point out is that most test takers were familiar with the formats of the task and item types on the reading section of the TOEFL iBT. According to participant background information, each respondent had taken the official test three times on average. This apparently helped test takers choose strategies selectively depending on the formats of the item types. Besides respondents' skill with the formats of the reading items, another possible explanation for the use of the most frequent strategies was that respondents were by and large highly proficient ESL learners. On average, test takers' self-reported ratings are High Intermediate for overall L2 proficiency and Intermediate for L2 reading proficiency. Therefore, it is plausible that test takers benefited from their relatively high level of proficiency in their choice of test-taking strategies that were mostly attuned to the requirements of the question items. This may also explain why most of the strategies that were used at the very high level of

frequency were test-management strategies (or construct-relevant response behaviors). This observation is in line with that of Cohen and Upton (2006) who had highly proficient ESL learners respond to reading tasks and items similar to those employed in this study. Cohen and Upton (2006) found that a major strategy trend among their respondents was the use of test-management strategies.

The second level of strategies comprised strategies that combined features of the highly and moderately frequent strategies. Thus, a portion of the middle-ranked strategies can be described as test-management and the other portion as test-wiseness. In other words, some of the strategies that were used at the high level of frequency involved the use of construct-relevant response behaviors, whereas others employed textual and/or technical aspects of the reading sets that were not construct-relevant. To illustrate with two strategies, some test takers used the strategy of option elimination with the BC-*ss* items while other test takers matched the content of the highlighted sentence and the options. While the former strategy required the test taker to check each option for its appropriateness as a replacement of the highlighted sentence, the latter was only limited to finding certain textual elements that are shared by the highlighted sentence and one of the options. However, for the above-mentioned reasons of test takers' skill and proficiency, the use of test management strategies at this level of frequency far exceeded that of the test-wiseness strategies.

The above description of highly frequent strategies is not to be taken to imply that all moderate-level strategies were test-wiseness. In fact, strategies with moderate levels of frequency also combined both test- management and test-wiseness strategies, but a look at strategies at this level across the item types would reveal that most of these

strategies were test-wiseness. The fact that test-wiseness characterized moderate-level strategies suggests that test takers managed well to answer the question items through the use of the first two levels of strategies that are mostly test-management. It also suggests that moderate-level strategies generally made up final steps of solving item types that test takers found to be challenging. This points out that test takers were able to make use of test-wiseness strategies in answering the various item types. Also, the nature of these strategies shows that test takers tended to follow systematic ways in using a variety of technical and textual elements of the test. Similar findings were reported by P. Yang (2000) who examined the impact of test-wiseness on performance on the TOEFL-CBT.

Strategies used by test takers in this study can be classified into two groups. The first group represents strategies that test takers used across several item types, and so these strategies possessed a high level of compatibility with a variety of item formats. Such strategies included eliminating certain options, deciding on an option upon hesitation among two or more options, confirming the selected answer, and reading the given paragraph either to locate the source of critical information or to determine the required information. These strategies occurred across the three task types; that is, the use of such strategies did not depend on whether a given item was basic comprehension, inference, or reading-to-learn. This does not preclude the fact that these strategies were used with some variation across item types, as is to be explained below. Noticeably, the strategy of option elimination was used with all of the item types. The use of this strategy was intended by test takers to limit the number of options they had to work with in their attempts to determine which option was the right one.

The strategy of confirming the selected answer is associated with nine item types including BC-*v*, BC-*fi*, BC-*nf*, BC-*pr*, BC-*ss*, I-*bi*, I-*it*, and I-*rp*. In other words, the strategy of answer confirmation was used with all items on the basic comprehension and inference tasks, which can be attributed to the fact that answers to the items that are part of these tasks can be confirmed more easily and directly compared with reading-to-learn task items. The strategy of paragraph reading for the purpose of locating the critical information or determining the required information occurred with five item types that include BC-*fi*, BC-*nf*, I-*bi*, I-*it*, and I-*rp*. Therefore, test takers had to read the given paragraph with two basic comprehension items and all of the three inference items. Obviously, test takers were able to distinguish the inference items from other task items and knew that they should read the paragraph with the inference items. The answers to such items require that test takers *read between the lines*, as mentioned by Participant (13).

The second group of strategies comprises those that were unique or specific to the item types with which they were used. This group of strategies covers virtually most of the strategies that occurred with very high frequency and some of the other strategies that appeared at the second and third levels of frequency. Examples of such strategies include reading the sentence containing the target word (BC-*v*1), attempting to locate each option in the paragraph while eliminating certain options (BC-*nf*1), reading the sentence containing the target pronoun (BC-*pr*1), reading the highlighted sentence while focusing on the main points (BC-*ss*8), trying the given sentence in one or more of the slots in the paragraph to see where it best fits (I-*it*4), selecting answer choices that represent important or summary points of the passage (R2L-*ps*6), and referring to the passage to

determine if a given geological process is “constructive” or “destructive” (R2L-*st*2). One notable difference between the common and the specific groups of strategies is that the specific strategies were more in harmony with the item format compared with the common strategies. The two strategy groups accord with what Rupp and his colleagues (2006) labeled as general strategies that can be applied to any test format, and item-related strategies that test takers use with question items.

Strategy use among test takers was purposeful, multi-form, and resourceful. Strategy use was purposeful in that test takers applied certain strategies for different goals. Such goals reflected distinct levels of language proficiency and test performance. For example, the strategy of reading the options first was used with five different item types (incl., BC-*v*, BC-*fi*, BC-*nf*, I-*bi*, and I-*rp*) in order to fulfill one or the other of three purposes: First, starting with the options first helped test takers become acquainted with the options. Second, when test takers thought they knew the answer from their background knowledge, they started with the options first in order to choose a preliminary option. And third, reading the options first was intended to help locate the critical information or determine the required information in the text.

The multi-form feature of strategy use manifests itself in the multiple variants a single strategy could take. Certain strategies were used by test takers with noticeable variation across several item types. A strategy like paragraph reading was used by test takers in different ways across five item types (incl., BC-*fi*, BC-*nf*, I-*bi*, I-*it*, and I-*rp*). With BC-*fi*, test takers read the paragraph or a relevant portion of it in order to locate the critical information, whereas with BC-*nf*, test takers read the whole paragraph or a portion of it to figure out the exception. Yet, with I-*bi*, test takers read the paragraph until

they could locate the critical information. Using variants of the same strategy across several item types suggests that these strategies can be tailored to fit various formats and demands of item types. Such strategies were applied with different frequency rates, presumably depending on how critical a given strategy was in answering a given item type. Thus, in order for test takers to infer the answer to an I-*bi* item, they had no choice but to read the paragraph. And, to determine the exception in response to BC-*nf,* test takers had to read either the whole paragraph or a portion of it since three out of the four options are scattered across the paragraph.

Strategy use among test takers was resourceful in that strategies were often applied using different means. Such means were largely determined by formats and demands of item types in addition to test takers' reading ability and test-taking skill. Hence, there were strategies that allowed test takers to experiment with a variety of ways to apply these strategies. Test takers made use of text-based information as well as certain technical elements when they wanted to confirm the answers they had chosen or decide on an option upon hesitation between two or more options. Thus, to decide on an option, test takers reread the given paragraph or a portion of it or looked for textual clues that would hint at the potential option. Similarly, confirming the selected answer involved rereading of the given paragraph or a portion of it, or checking options other than the chosen one. It is interesting to see that the two strategies of deciding on an option and confirming the selected answer shared the use of one aid, namely rereading the given paragraph or a portion of it. It seems that test takers trusted paragraph reading as the key resource of the information they needed to end up with the right option.

Taken together, the three features of strategy use in this study point out that test takers enjoyed a high ability to, first, choose test-taking strategies to serve various goals of answering different item types; second, use distinct forms of strategies to deal with the diverse item formats; and third, employ strategies using several means.

Research question two.

Are there any differences between high- and low-scorers among subjects in their use of test-taking strategies on the reading tasks and items? In order to address the second research question, high- and low-scorers among test takers were identified on the basis of their total scores on the reading sets. High scorers had the top five scores (≥ + 0.95 standard deviation), whereas low scorers had the bottom five scores (≤ − 0.79 standard deviation). The decision as to who among test takers would fit in which group was endorsed by the results of item analyses which helped assign ranks to test takers according to their performance on the question items. In order to maintain the secrecy of the real names of participants in the high- and low- scoring groups, an anonym is assigned to each participant for the purpose of data analysis and reporting.

Intercorrelations of participants' total and reading scores on their last TOEFL iBT and their total scores on the reading sets in this study were calculated. Substantial positive correlations were shown between participants' TOEFL-iBT total scores and scores on the reading sets: r (23) = 0.651, p<.01, and participants' TOEFL-iBT reading scores and scores on the reading sets: r (23) = 0.548, p<.01. These correlation coefficients support the researcher's earlier predictions that participants' levels of overall proficiency and reading proficiency would correlate with their scoring levels in this study as determined by the reading sets. They also provide us with a rough indication that

participants responded to the reading sets to the best of their overall proficiency and reading ability in English. Table 16 below displays high- and low-scorers' raw and standard or z-scores, anonyms, and ranks in the main group as indicated by the results of item analyses.

Table 16
High- and Low-scorers' Raw and Standard or Z-scores, Anonyms, and Ranks

High Scorers				Low Scorers			
Raw Score	Z- Score	Anonym	Rank	Raw Score	Z- Score	Anonym	Rank
25	+1.72	Ahmad	1	12	−0.79	Saja	21
24	+1.52	Badrah	2	10	−1.17	Tariq	22
22	+1.14	Deeb	3	10	−1.17	Waleed	23
22	+1.14	Emad	4	8	−1.56	Yasser	24
21	+0.95	Fahdah	5	4	−2.33	Ziyad	25

Note. Standard deviation of test takers' total scores is 5.20.

Comparisons between high- and low-scoring groups made use of frequency counts. Differences in frequency counts of strategy use can point to discrepancies between high- and low-level writers or readers (Kasper, 1998). In addition, as Kasper (1998) recommended, frequency counts in this analysis were supplemented with details about the manner in which strategies were used differentially by the two discrete levels of scoring and performance among participants. It was assumed that strategies with differences of two or more frequency counts would prove noteworthy to report in answer to the second research question. For item types that have two instances across the reading sets, there were a few strategies whose frequency counts exceeded five for a given scoring-group. Comparisons of strategy use by the two groups are presented here by task and item type, and bar charts are used to display differences in strategy frequencies between the two groups. Strategy descriptions are included along with the comparisons.

Strategy abbreviation(s) specific to this section:

- (TT-*i*#): Task type, item type, and strategy serial number, respectively. (E.g., BC-*v*3 denotes strategy 3 among strategies used with Basic Comprehension-*vocabulary items* or BC-*v* items).
- (Anonym, RS#I#): Participant's anonym, serial number of the reading set (1 for Reading Set 1, and 2 for Reading Set 2), item number, respectively.

***High- versus low-scorers' strategy use across task items*.**

Basic comprehension-vocabulary (BC-v). As can be seen in Figure 3 below, the high- and low-scoring groups were different in their use of five strategies. Among all strategies used with the two BC-*v* items as shown in Table 6, the strategies that appear in Figure 3 were the ones associated with more notable differences between the two groups.

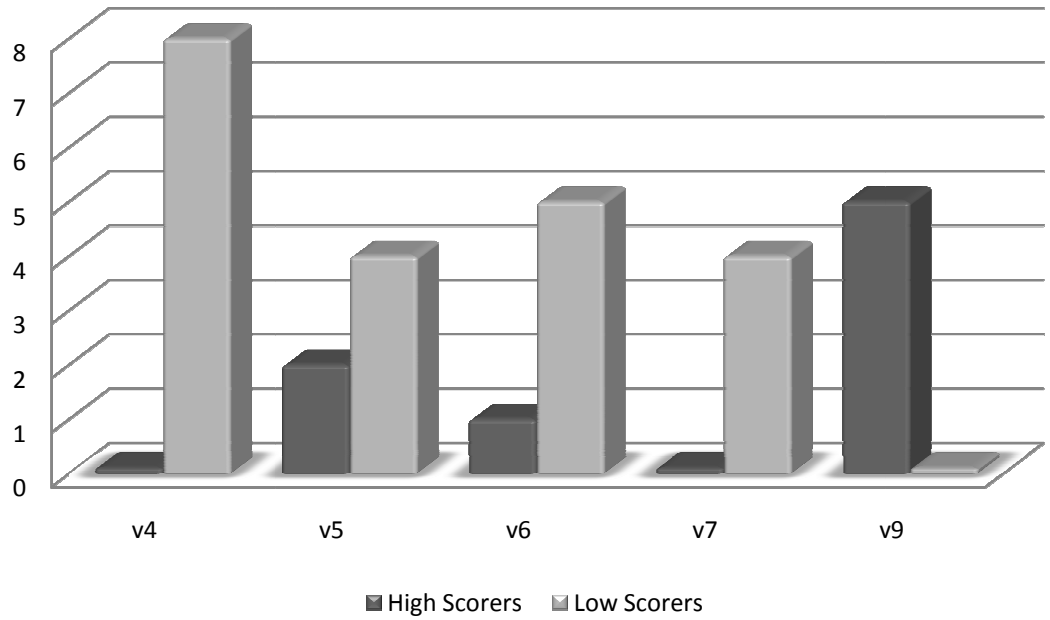

Figure 3. Strategy frequencies of high scorers versus low scorers on BC-*v* items.

Clearly, low scorers surpassed their high counterparts in their use of four out of the five strategies in Figure 3. To begin with, low scorers tended to read the context surrounding the sentence that contains the target word besides reading the sentence itself (*v*4).

> *I knew two or three options out of the four options … I had to read the sentence before the one containing the word, the sentence itself, and the next sentence to understand what the author says* <*v*4>. (Waleed, RS2I4)

Here Waleed used strategy (*v*4) because he needed to understand what the author meant by the target word. Therefore, reading the sentence with the target word by itself did not seem to have helped him figure out the meaning of the target word.

Another strategy used more frequently by low scorers was replacing the target word with each one of the options to see which option would be the best fit in context (*v*5).

> *First, I read the options to familiarize myself with them. Then, I read the sentences before and after the sentence containing the word, besides the sentence itself, in order to understand the context of the word* <*v*4>. *I also tried the four options in place of the* [target] *word* <*v*5>. (Ziyad, RS1I1)

Ziyad made use of strategy (*v*5) after using strategy (*v*4). This indicates that Ziyad could not figure out the meaning of the target word by applying strategy (*v*4), so he resorted to strategy (*v*5).

Low scorers made more frequent use of the strategy of option elimination (*v*6). This mirrors the purpose of using this strategy among test takers in general; that is, low scorers aimed to reduce the number of alternatives they had to deal with on the BC-*v* items: *I did not know "relics." I knew only Options 2 "regions" and 3 "remains." … I eliminated the other two options* [1 and 4] *from the beginning* <*v*6> (Yasser, RS2I4). The reason why Yasser eliminated two options on the given item was because these options

had words that were not familiar to him. Consequently, he kept for consideration the options with the vocabulary known to him.

Low scorers showed more reliance on guessing (*v*7). More often, guessing for low scorers was the last step in the process of determining which option was the right answer.

> *Here, this word puzzled me, 'honestly.' I chose Option 2 "regions" based on the idea of the paragraph, which is a discussion of mountains and locations and heights like "the Himalayas" and "the Caledonians;" so I guessed the answer* <*v*7>. (Tariq, RS2I4)

Tariq had to guess the answer rationally here because he was puzzled by the target word. He based his guessing on the central idea of the paragraph, namely geographical manifestations such as mountains and heights. Thus, in Tariq's view, "regions" is the closest among all options to this idea.

In contrast to the low-scoring group, high scorers made more use of background knowledge on BC-*v* in order to think of the word meaning (*v*9): *I knew that 'measurably' means something that 'can be measured'* <*v*9>*, and so "immeasurably" means something you cannot measure* (Badrah, RS1I1). Badrah retrieved the meaning of the antonym of the target word from her background knowledge so that she could assign the right meaning to the target word. High scorers also used background knowledge to choose an option (*v*9): *I chose Option 3 "remains." I used my 'background knowledge' for the meaning of the word* <*v*9> (Fahdah, RS2I4). Fahdah made a straight choice of Option 3 using her background knowledge.

Basic comprehension-factual information (BC-fi). As can be seen in Figure 4 below, the high- and low-scoring groups were different in their use of three strategies. Among all strategies used with the two BC-*fi* items as shown in Table 7, the strategies

that appear in Figure 4 were the ones associated with more notable differences between the two groups.

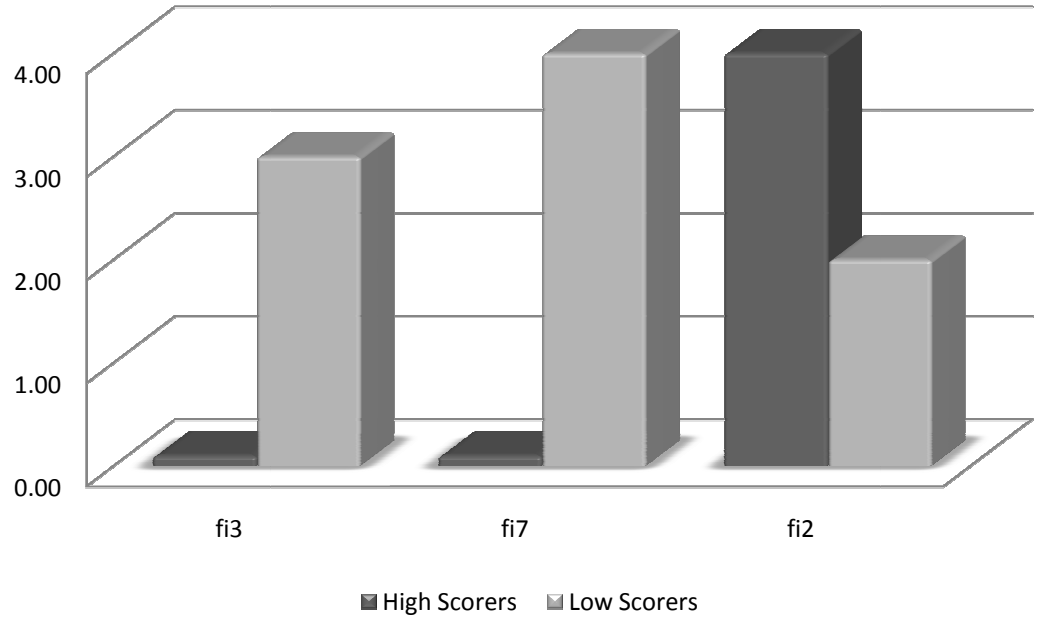

Figure 4. Strategy frequencies of high scorers versus low scorers on BC-*fi* items.

According to Figure 4, low scorers used two strategies more frequently than did high scorers. First, low scorers chose to reread the paragraph referred to by the BC-*fi* item or a portion of it (*fi*3): *I reread the part of the paragraph around the* [target] *word in the last item* [Item 6: BC-*v*] <*fi*3> to *see what the paragraph talks about regarding the government* (Ziyad, RS1I7). Ziyad was familiar with the source of critical information in the paragraph because he read the paragraph while he was answering the previous item. With this item, he thought that he could just reread this portion of the paragraph in order to determine the required information.

Another strategy used more frequently by low scorers was matching key word(s) or phrase(s) in the item stem or options and the paragraph (*fi*7): *I skimmed the paragraph fast starting from the point of stop in the previous item "Government's role …", trying to*

understand. I matched a key word "religion" in the paragraph with an option [Option 1: "Creating a state religion"] <*fi*7> (Yasser, RS1I7). Yasser resorted to matching the information in the options and the paragraph. As a result, he picked an option using a key word that is shared by Option 1 and the paragraph.

On the other hand, high scorers made more use of the strategy of confirming the selected answer (*fi*2).

> *I read the paragraph carefully, paying attention to as many specific details as possible to know 'the things that they used to support the government with or their role or effect.' I read the options and chose Option 3, and reread the paragraph to confirm my answer* <*fi*2>. (Fahdah, RS1I7)

Although Fahdah read the paragraph carefully at first, she chose to reread it in order to confirm the answer she had chosen.

Basic comprehension-negative fact (BC-nf). As can be seen in Figure 5 below, the high- and low-scoring groups were different in their use of three strategies.

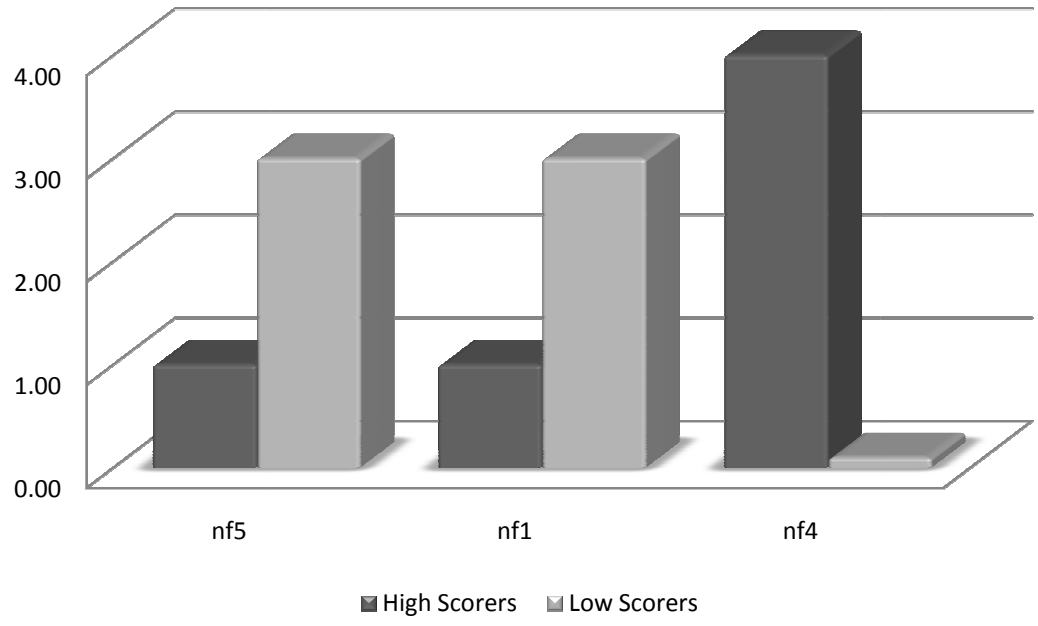

Figure 5. Strategy frequencies of high scorers versus low scorers on BC-*nf* item.

Among all strategies used with the BC-*nf* item as shown in Table 8, the strategies that appear in Figure 5 were the ones associated with more notable differences between the two groups. According to Figure 5, low scorers used strategy (*nf*5) more frequently than did high scorers. When using strategy (*nf*5), low scorers preferred to read the options before the paragraph: *I read the options <nf*5*>. I scanned the paragraph for the options one by one, and used elimination of options to determine the answer* (Ziyad, RS1I10). Ziyad read the options before the paragraph in order to have familiarity with them so that it would be easier for him to locate such options in the paragraph.

Low scorers preferred to locate each option in the paragraph and eliminate it if it existed (*nf*1).

> *I love EXCEPT items though they take time. I scanned the paragraph trying to locate each one of the options. I found "workers …"* [Option 1]*, "planters …"* [Option 2] *<nf*1*> … The only option that I did not find was Option 3.* (Saja, RS1I10)

Saja enjoyed working with the BC-*nf* item because all she needed to do was to identify which options were mentioned. By doing so, she chose the option that was left as her answer.

On the other hand, high scorers used strategy (*nf*4) more frequently than did low scorers. Using strategy (*nf*4), high scorers read or skimmed the paragraph while eliminating certain options.

> *I read the paragraph fast and found that "Democrats attracted farmers isolated from the market … or uncomfortable with it," so thought that this is something* [Option 2] *I cannot chose <nf*4*>. Also, Option 3 is mentioned <nf*4*>. Also, Option 4 is mentioned <nf*4*>. That is why I chose the option that remained* [Option 1]. (Badrah, RS1I10)

Badrah started her response to this item by reading the paragraph and at the same time she was trying to determine which options she should eliminate. Therefore, she

eliminated three out of the four options because she found them in the paragraph, and chose the one that was not mentioned.

Basic comprehension-pronoun reference (BC-pr). As can be seen in Figure 6 below, the high- and low-scoring groups were different in their use of three strategies. Among all strategies used with the BC-*pr* item as shown in Table 9, the strategies that appear in Figure 6 were the ones associated with more notable differences between the two groups.

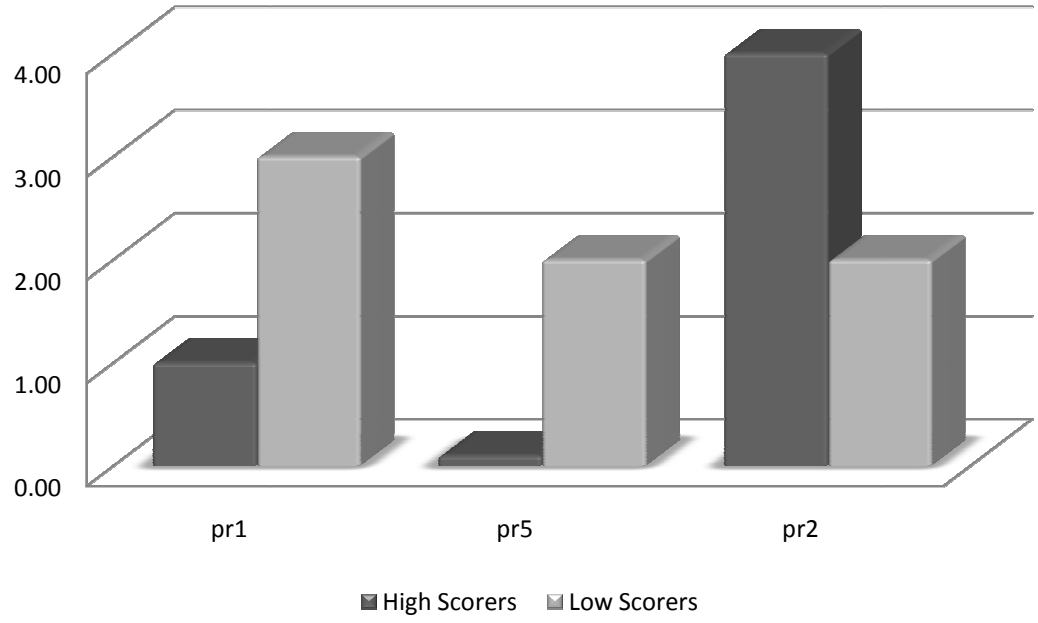

Figure 6. Strategy frequencies of high scorers versus low scorers on BC-*pr* item.

According to Figure 6, low scorers used two strategies more frequently than did high scorers. First, low scorers assigned more preference to reading the sentence that had the pronoun to identify the referent (*pr*1): *I read the sentence from the beginning <pr*1*>. I stopped because there are unknown vocabs. So, I had to read the sentence more than once. I do not know "valleys" and "slowly moving masses of ice"* (Waleed, RS2I8). Waleed reported that he read the whole sentence with the target pronoun. The frequent

use of strategy (*pr*1) among low scorers indicates that they generally found BC-*pr* challenging, as is clear from Waleed's use of a rereading strategy.

Low scorers were inclined to take advantage of clues in their choice of an option (*pr*5). This serves as an indication that low scorers not only attempted to read and reread the sentence containing the pronoun, but also made use of clues that could help them get the BC-*pr* item right: *I saw "carrying" and "valleys," and thought that 'nothing can carry huge quantities of eroded rock debris except valleys.' Therefore, I chose Option 3 "valleys"* <*pr*5> (Yasser, RS2I8). Yasser did not manage to understand the sentence containing the pronoun to the extent that would enable him to determine an option to choose. Therefore, he thought about relating elements of the content (e.g., "carrying", "valleys", "huge quantities of eroded rock debris") to one another semantically. As a result, he chose Option 3 "valleys".

High scorers, on the other hand, had a preference for confirming their answers (*pr*2) before moving on to the next item.

> *Most likely, it* [the pronoun] *refers to "masses of ice"* [Option 2]. *I chose Option 2 using the context. It is impossible that the answer is one of the other options. I reread* [the sentence] *to confirm my answer* <*pr*2>. (Ahmad, RS2I8)

Ahmad was almost certain that Option 2 was the best option available to stand for the referent. Even so, he preferred to reread the sentence containing the pronoun to confirm that he picked the right choice.

Basic comprehension-sentence simplification (BC-ss). As can be seen in Figure 7 below, the high- and low-scoring groups were different in their use of five strategies. Among all strategies used with the two BC-*ss* items as shown in Table 10, the strategies

that appear in Figure 7 were the ones associated with more notable differences between the two groups.

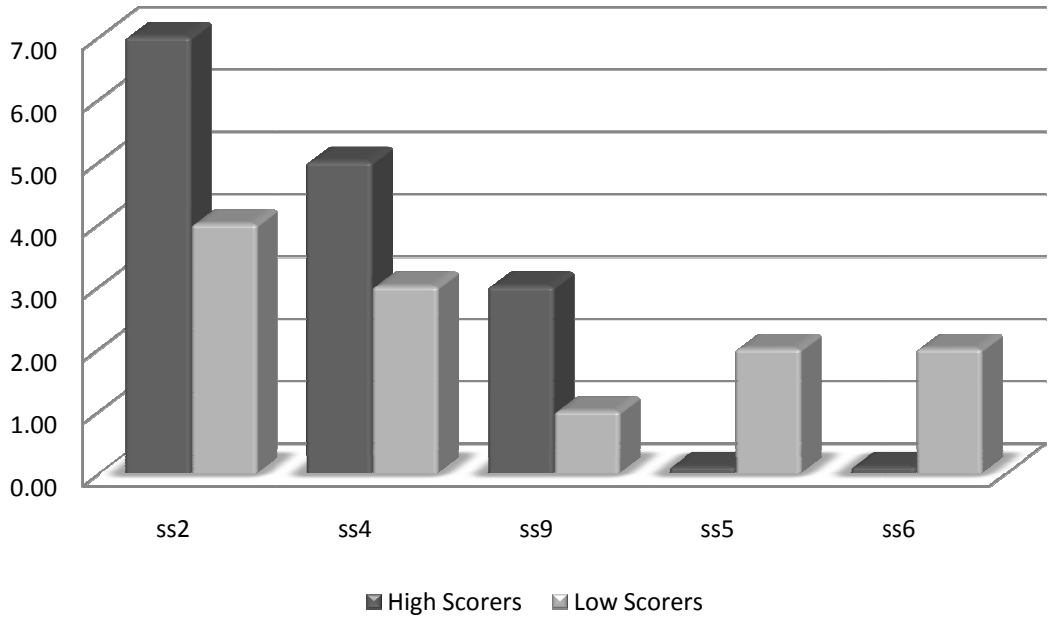

Figure 7. Strategy frequencies of high scorers versus low scorers on BC-*ss* items.

According to Figure 7, high scorers used three strategies more frequently than did low scorers. First, high scorers were more disposed to use clues to choose an option (*ss*2), especially when they were not able to determine the answer using understanding of the highlighted sentence.

> *I read the options, they were close. I considered Option 2 which contains "although" because of "but" in the highlighted sentence; "although" is a synonym of "but" <ss2>. So, the meaning of the highlighted sentence and the meaning of Option 2 are very close … so chose Option 2 "Although they seem permanent …."* (Deeb, RS2I9)

Deeb chose Option 2 on the basis of a discoursal element the option shares with the highlighted sentence, specifically the meaning of the conjunctions *although* and *but*.

The second strategy that was more frequent among high scorers was elimination of options (*ss*4): *I read the highlighted sentence, understood it generally, and matched its*

general meaning with those of the options. Option 1 was not correct <*ss*4>. *Option 2 was closer to the highlighted sentence, so I chose it directly* (Fahdah, RS2I9). Fahdah used her global understanding of the highlighted sentence to determine which options to eliminate and which one to choose as her answer. Accordingly, she eliminated the options that did not conform with the highlighted sentence in meaning.

Strategy (*ss*9) was among the strategies that high scorers used more frequently. Using strategy (*ss*9), high scorers made more use of confirmation of their answers to BC-*ss*.

> *"Although" in Option 2 is close to "but" in the paragraph* [highlighted sentence], *so I chose Option 2 on this basis … I read the other options and found none of them to give the same meaning as that of the highlighted sentence; none is as close to the highlighted sentence in meaning as Option 2* <*ss*9>. (Emad, RS2I9)

Emad confirmed his answer by reading the other options to see if any of them would be better than the option he had chosen on the basis of the meaning of the highlighted sentence.

On the other hand, low scorers made more frequent use of two strategies than did high scores. Low scorers experienced more hesitation as to which option to choose; therefore, they had to decide on an option (*ss*5).

> *The difference among the options is slim. I eliminated Options 3 and 4 because they say something the author does not talk about. Option 1 gives a specific time* <*ss*5>, *Option 2 does not give a specific time. I expected it* [Option 1] *to be wrong, I do not know.* (Ziyad, RS2I9)

Ziyad found barely any difference among the options. He tried the elimination strategy, and he could eliminate two options because these options sounded to him irrelevant to what the author said. He felt hesitant between the two other options, but decided on one of them because it mentioned a definite time as did the highlighted sentence.

Also, low scorers had to read the highlighted sentence more than one time to get as much grasp of its meaning as possible (*ss*6). *I read the highlighted sentence more than one time, trying to understand it <ss*6*>. I understood it though it has some unknown vocabulary (e.g., "epitome of permanence")* (Waleed, RS2I9). Waleed had to read the highlighted sentence in the given item more than one time until he understood it.

Inferencing-basic inference (I-bi). As can be seen in Figure 8 below, the high- and low-scoring groups were different in their use of four strategies. Among all strategies used with the two I-*bi* items as shown in Table 11, the strategies that appear in Figure 8 were the ones associated with more notable differences between the two groups.

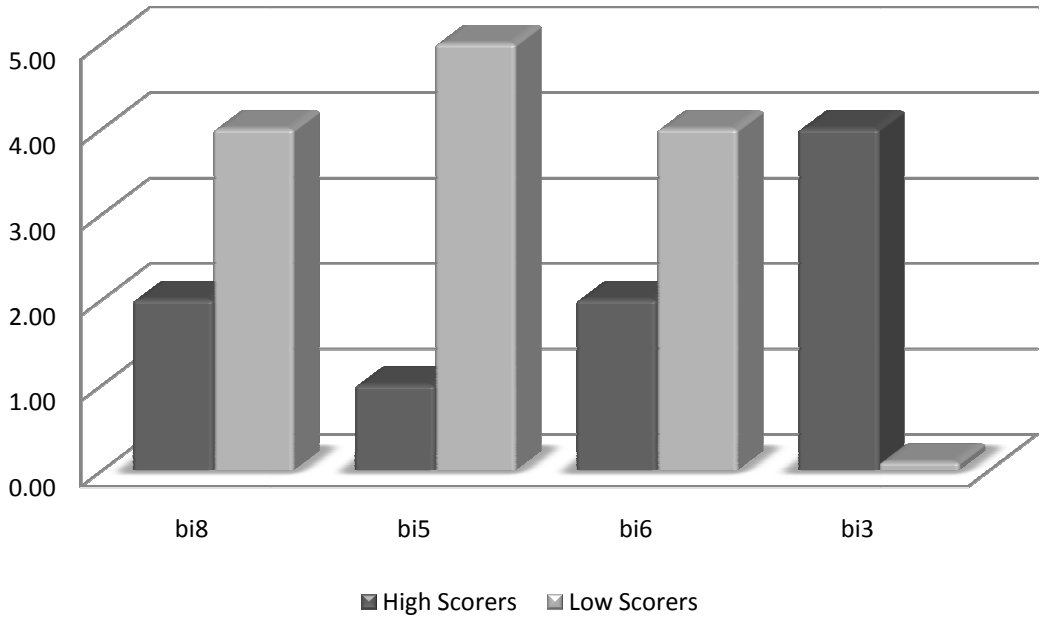

Figure 8. Strategy frequencies of high scorers versus low scorers on I-*bi* items.

According to Figure 8, low scorers used three strategies more frequently than did high scorers. First, low scorers were more inclined to confirm their answers to the I-*bi* items (*bi*8).

> *Option 1 was the only one that makes sense. I read here* [pointing to the relevant portion of the paragraph] *that "the Himalayas" are only "50 million years old."*

> *Therefore, the height of the mountain is not a yardstick to determine its age. I checked the other options to confirm my answer <bi*8>. (Tariq, RS2I3)

Tariq used his understanding of the critical information in the paragraph to choose the option he thought would most likely be the right one. Then, he checked the other options to make certain that none of these options would rival the option he had chosen.

Low scorers benefited from the use of clues or guessing in their choice of an option (*bi*5), especially when it was difficult to determine the answer on the basis of understanding of the paragraph.

> *I just felt desperate to answer it* [this item]. *I read "individual" in the paragraph and on the basis of which I chose Option 2 "… members to leave the Whig party" <bi*5>*; it suggests some division among the Whigs.* (Saja, RS1I9)

Saja chose her answer to this item using a semantic clue, that is, by relating a key word (i.e., "individual") in the paragraph and the content of Option 2.

Low scorers made more frequent use of key words or phrases in their attempts to locate the critical information in the paragraph (*bi*6).

> *I took a quick look at the options, trying to find any clue that can help me with the answer. Then, I skimmed the whole paragraph. I read carefully when I saw the word "Himalayas" <bi*6>*. I chose Option 1 because I deduced this information* [in Option 1] *from the general meaning of the paragraph.* (Waleed, RS2I3)

Waleed used his reading of the options to note key words that could help him find the answer in the paragraph. The key word he used for this purpose was "Himalayas".

On the other hand, high scorers used one strategy more frequently than did low scorers. Using strategy (*bi*3), high scorers reread the paragraph or a portion of it.

> *I chose Option 1, I thought about Option 4 " ... regional interests," but thought that these would follow certain geographic locations or places and I referred to the paragraph to check* [Option 4] <*bi*3>*, but did not find information that supports this option.* (Badrah, RS1I9)

Badrah chose Option 1, but then she thought that Option 4 was likely to be the answer. Therefore, she reread the portion of the paragraph that had information pertinent to Option 4 in order to check its adequacy. Eventually, she decided to keep Option 1 as her choice.

Inferencing-insert text (I-it). As can be seen in Figure 9 below, the high- and low-scoring groups were different in their use of seven strategies. Among all strategies used with the two I-*it* items as shown in Table 12, the strategies that appear in Figure 9 were the ones associated with more notable differences between the two groups.

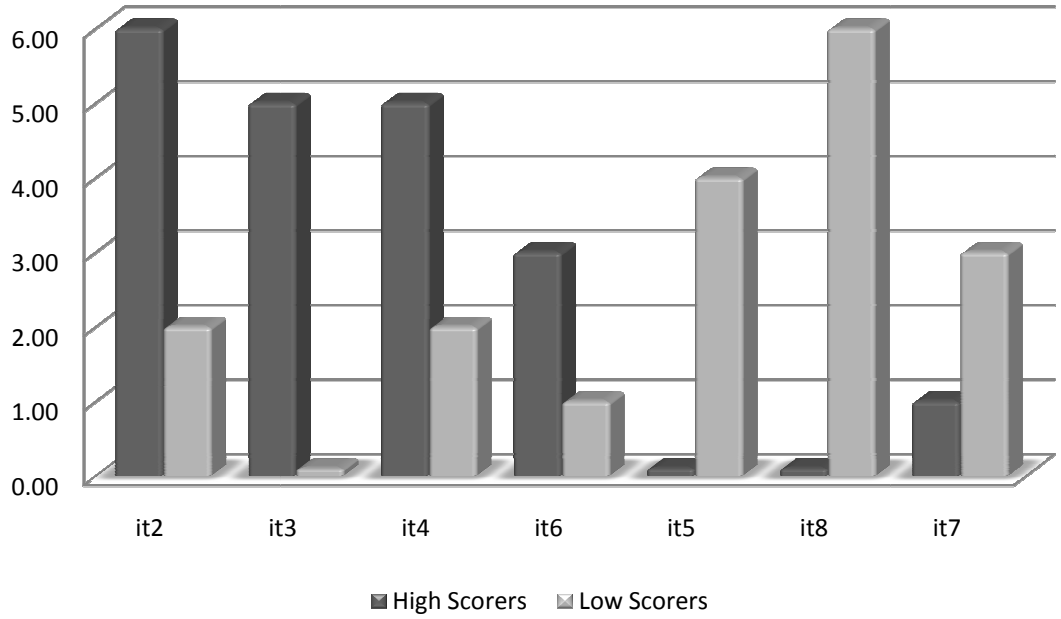

Figure 9. Strategy frequencies of high scorers versus low scorers on I-*it* items.

According to Figure 9, high scorers used strategies (*it*2, *it*3, *it*4, and *it*6) more frequently than did low scorers. First, high scorers exhibited a tendency to read the paragraph in order to determine in which slot the given sentence would best fit (*it2*).

> *I read the paragraph till I got to number 4* [Slot 4]. *I noticed that number 1* [Slot 1] *best fits as a beginning <it*2*>. The previous sentence talks about forming "the Whig party," so the next one would start with "The new party … ."* (Ahmad, RS1I12)

Ahmad read the paragraph referred to by the given I-*it* item from the beginning. He found that the first sentence in the paragraph introduced the Whig party, so he thought that the given sentence could be inserted just following the first sentence. Thus, the given sentence would give more information or a supporting detail about “the Whig party”.

High scorers chose the option representing the slot where appeared to be a logical connection between the given sentence and the sentence(s) before and/or after the given slot (*it*3).

> *First, I read the given sentence. I could tell how the previous sentence would sound like through the first three words "This new party" in the given sentence. I tried to find a sentence with relevant information* [about “This new party”] *before each slot. I chose Option 1* [Slot 1] *because the sentence before it ends with "to form the Whig party," and then the author would be expected to start the next sentence with "This new party" which refers to the party that was formed* [the “Whig party”] <*it*3>. (Deeb, RS1I12)

Because the given sentence began with the phrase “This new party”, Deeb tried to locate a sentence that said something about that phrase. And, because the first sentence in the paragraph suggested that a new party was formed near the end of it, Deeb felt certain that the given sentence must follow the first sentence of the paragraph.

High scorers also had a tendency to try the given sentence in one or more of the slots in the paragraph to determine in which slot it best fitted (*it*4): *I tried the given sentence in each one of the places* [slots] *to see where "Under different climatic conditions"* [the given sentence] *can best fit* <*it*4>. *I chose Option 1* (Emad, RS2I11). Emad used strategy (*it*4) which seemed to have helped him pick an answer by just reading the first few sentences of the paragraph.

Confirmation of the selected option was another frequent strategy among high scorers (*it*6).

> *I found that the given sentence fits in Place 1* [Slot 1] *because the first sentence says "… to form the Whig party" and the given sentence begins with "This new party …," so I chose Option 1. I checked the other options, but felt that Option 1 is the best* <*it*6>. (Badrah, RS1I12)

Though I-*it* has its unique format, it lends itself to the strategy of confirming the selected answer. Badrah applied a confirmation strategy by checking the other slots or options.

On the other hand, low scorers used strategies (*it*5, *it*8, and *it*7) more frequently than did their high-scoring counterparts. The strategy of eliminating certain options (*it*5) was used more frequently among low scorers.

> *I eliminated Options 1 and 2, because if the given sentence is inserted in any of these places* [slots], *the meaning of the context will be messed up* <*it*5>. *This is because it says "another type" in the given sentence, so there must be a certain type in order for another type to be mentioned like "the wind" and "tree roots," so it has to be either Option 3 or 4.* (Tariq, RS2I11)

Tariq eliminated two options because neither of them identified a context that had the first type of weather conditions under discussion.

Low scorers made use of clues or guessing (*it*8) when they could not answer the I-*it* item through understanding.

> *"Even living things …." The word "even" is related to something that is mentioned before. Then, the author talks about "Tree roots" in the next sentence, and the given sentence says "Under different climatic conditions, another type of destructive force contributes to erosion." I was certain that the answer is Option 4 because the author here talks about 'another type' and starts talking about "Tree roots," so I chose Option 4* <*it*8>. (Yasser, RS2I11)

Yasser guessed the answer rationally, seeing that the phrase "tree roots" represented the other type the author referred to in the given sentence. Yasser understood "even" in the sentence before the one starting with "Tree roots" to be related to the first type of climatic conditions that cause erosion.

The third frequent strategy among low scorers was deciding on an option (*it*7) upon hesitation.

> *I was going to choose Option 2 because it talks about a second point on "erosion," so I thought that the previous sentence must talk about "erosion." Therefore, the best potential place is Option 1, but Option 2 is also the same thing and it seemed to me that the second sentence is a completion of the first one, and so the given sentence cannot come in between the two. I chose Option 3 because "living things" also affect* <*it*7>. (Saja, RS2I11)

Saja hesitated as to which one among the three options she should choose. She decided on Option 3, thinking that "living things" play a role in erosion.

Inferencing-rhetorical purpose (I-rp). As can be seen in Figure 10 below, the high- and low-scoring groups were different in their use of five strategies. Among all strategies used with the two I-*rp* items as shown in Table 13, the strategies that appear in Figure 10 were the ones associated with more notable differences between the two groups.

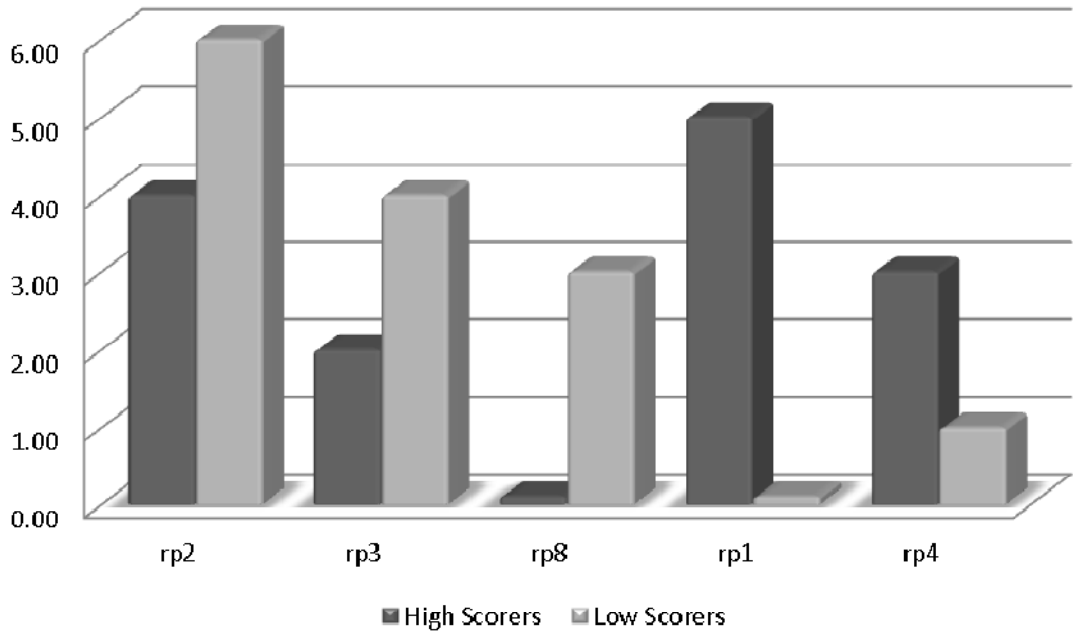

Figure 10. Strategy frequencies of high scorers versus low scorers on I-*rp* items.

According to Figure 10, low scorers used strategies (*rp*2, *rp*3, and *rp*8) more frequently than did high scorers. First, low scorers had a tendency to read the sentence containing the highlighted phrase and the surrounding context in order to figure out the required information (*rp*2): *Item 3* [I-*rp* #3 on Reading Set 1] *talks about "bankers and investors" in Paragraph 2. I read the sentence that has the highlighted phrase and the sentence after it <rp2> to understand what the topic is about* (Tariq, RS1I3). Tariq read the sentence with the highlighted phrase and the next sentence, thinking that this would help him infer the required information.

Another frequent strategy among low scorers was reading the paragraph the I-*rp* item referred to or a portion of it more than once (*rp*3).

> *I read the paragraph from the beginning more than one time <rp3>, and understood "the main agent of erosion," but 'I do not know the meaning of erosion,' but understood that "carbon dioxide" is 'one of the causes or results of erosion.'* (Waleed, RS2I6)

Waleed had to read the paragraph referred to by the given I-*rp* item more than once so that he could reach an adequate level of understanding that would enable him to infer the required information.

Low scorers resorted to matching the content of the sentence containing the highlighted phrase and the options (*rp*8).

> *I read the sentence before the one with the highlighted phrase, the sentence with the highlighted phrase, and the one after it. I found key words, and tried to locate matching information in the options <rp8>… I chose Option 2 "unfairly becoming rich" because of "unearned wealth" in the paragraph.* (Saja, RS1I3)

Saja read the content of the highlighted sentence and its surrounding context. Then, she read the options while trying to locate information that matched the content of the portion

of the paragraph she read. She identified Option 2 which had a phrase that almost meant the same thing as a phrase mentioned in the paragraph.

On the other hand, high scorers used two strategies (*rp*1 and *rp*4) more frequently than did their low-scoring counterparts. High scorers were more disposed to read the options first (*rp*1): *First, I skimmed the options to know what options are available <rp*1*>so that I could return to them later* (Fahdah, RS1I3). Fahdah skimmed the options first so as to get herself acquainted with them before reading the text. Fahdah also mentioned that she would turn to the options either while or after reading the text.

High scorers also tended to confirm their answers (*rp*4): *I read the options first. Of course, I read the paragraph and found the answer to be this* [Option 1]. *And, I returned to the paragraph to make sure I chose the right answer <rp*4*>* (Deeb, RS2I6). Deeb confirmed his answer to the given I-*rp* item by referring to the paragraph.

Reading to learn-prose summary (R2L-ps). As can be seen in Figure 11 below, the high- and low-scoring groups were different in their use of four strategies. Among all strategies used with the R2L-*ps* item as shown in Table 14, the strategies that appear in Figure 11 were the ones associated with more notable differences between the two groups.

According to Figure 11, high scorers used strategies (*ps*6, *ps*3 and *ps*1) more frequently than did low scorers. Using strategy (*ps*6), high scorers selected answer choices that represented important or summary points of the passage.

> *I read all the answer choices first fast. I tried to think about what choices are more prominent or more important <ps*6*>. I found that Choice 5 'talks about the difference between the Whigs and the Democrats parties,' so I selected it … I selected Choice 6 … Then, I selected Choice 4 'because it also expresses another difference between the Democrats and the Whigs.'* (Badrah, RS1I13)

Badrah read the answer choices one by one. Upon reading each answer choice, she asked herself if it was one of the most salient or important points discussed in the passage. Accordingly, she selected all three answer choices required by R2L-*ps*.

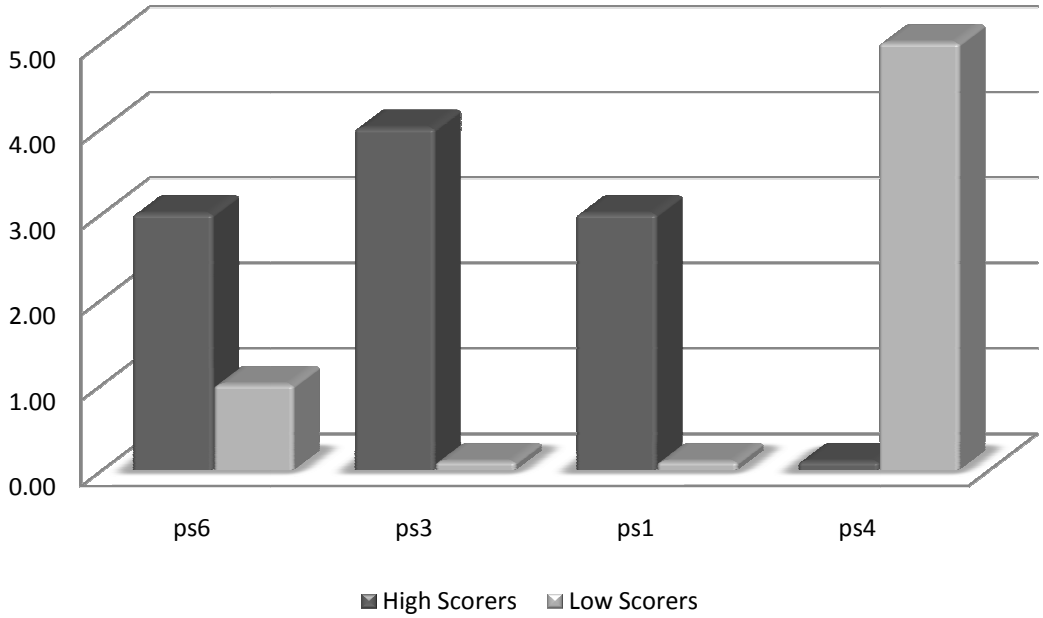

Figure 11. Strategy frequencies of high scorers versus low scorers on R2L-*ps* item.

High scorers also made more elimination of answer choices (*ps*3).

> *I had one minute with this item. I tried to read all the choices, but it was difficult. I selected Choice 1, then I read it, and it sounded OK to me. I read Choice 2, but it did not suit me <ps3>. I selected Choice 3. I looked at the other choices, but they do not sound suitable <ps3>.* (Deeb, RS1I13)

Deeb managed to give due consideration to all answer choices on this item though he was under time pressure. Using strategy (*ps*3), Deeb was able to determine which answer choices to select by reading them only one time.

High scorers used the understanding they had developed as a result of skimming or reading the passage to select answer choices (*ps*1): *I chose Choice 2 "During Andrew Jackson's two terms as President …." I remembered from my reading of the passage that*

'during his presidency, a new party was formed' <*ps*1> (Fahdah, RS1I13). Fahdah drew on her reading of the passage to decide if she could pick Choice 2.

On the other hand, low scorers used only one strategy more frequently than did high scorers. This strategy was the selection of answer choices by means of random guessing (*ps*4): *20 seconds left … I selected Choice 2 because I managed to read it and I understood that it can be a summary point. I selected the two other choices* [Answer Choices 4 and 5] *randomly* <*ps*4> (Tariq, RS1I13). Tariq worked on this item under time pressure, which explains why he selected two out of three answer choices using random guessing.

Reading to learn-schematic table (R2L-st). As can be seen in Figure 12 below, the high- and low-scoring groups were different in their use of five strategies.

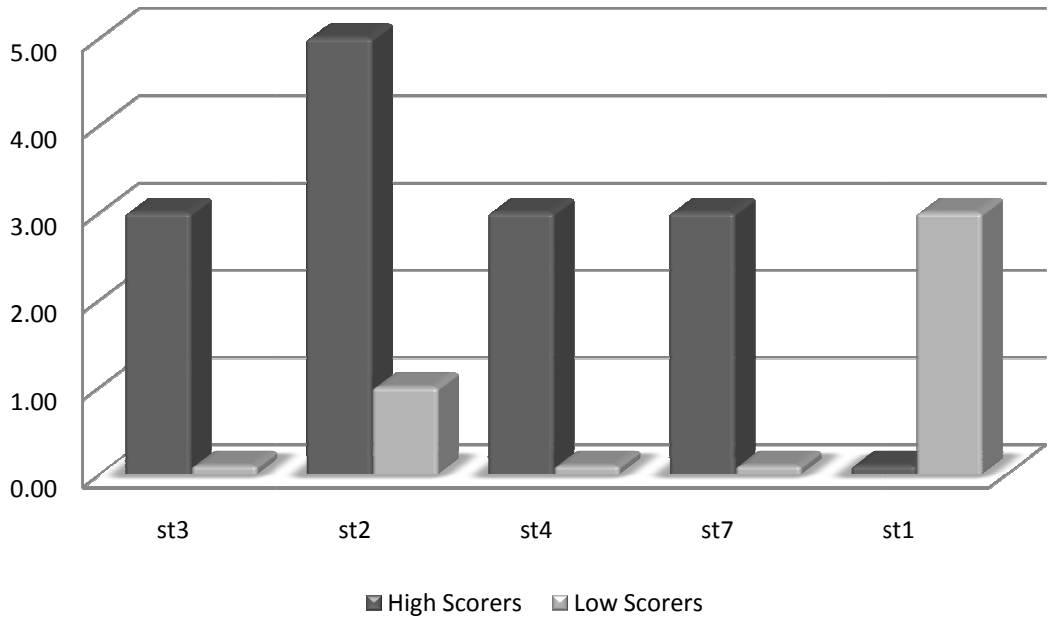

Figure 12. Strategy frequencies of high scorers versus low scorers on R2L-*st* item.

Among all strategies used with the R2L-*st* item as shown in Table 15, the strategies that appear in Figure 12 were the ones associated with more notable differences

between the two groups. According to Figure 12, high scorers used strategies (*st*3, *st*2, *st*4 and *st*7) more frequently than did low scorers. Using strategy (*st*3), high scorers used the understanding they had developed as a result of going through the previous items and reading their respective paragraphs to select answer choices.

> *This item is better and easier than the one before* [R2L-*ps*], *because it has short phrases in contrast to complete sentences* [in R2L-*ps*] … *My selections here were based on my answers* [to the previous items] <*st*3>. (Deeb, RS2I12)

Deeb used the information he had learned from answering the previous items to help him select answer choices on this item.

High scorers tended to consult the passage to check certain answer choices when they felt uncertain about whether such answer choices represented "constructive" or "destructive" processes (*st*2).

> *I read the answer choices and the two tables* [headings of the table], *and started selecting choices. Because I still had some time, I went to the passage and read a little bit <st2>, and then returned to the items. I selected the other answer choices and made some replacements or changes when I felt not certain about the answer choices I had selected.* (Fahdah, RS2I12)

Fahdah referred to the passage to read some information pertinent to the answer choices. By doing so, she was able to check the answer choices she had not selected yet.

High scorers also preferred to confirm their selections (*st*4).

> *I selected Choice 1 as "constructive" because this is what forms landscapes … Then I selected Choice 5 as "destructive" though I was not sure if it is mentioned in the passage or not… Also, I was not sure if Choice 2 is mentioned or not … I selected Choice 6 as "destructive" because of "rock debris" and "hot springs." I thought that this choice is mentioned in the passage either explicitly or implicitly … I returned to the passage to confirm my selections <st*4>. (Emad, RS2I12)

After Emad had selected most of the answer choices, he referred to the passage to make sure that his selections were the right ones.

High scorers also used the understanding they had developed as a result of skimming or reading the passage to select answer choices (*st*7): *I gained most of my understanding which helped me answer this item from my first reading of the passage* <*st*7> (Fahdah, RS2I12). Fahdah provided this report in an answer to the researchers' question about whether she used her reading of the passage or dealing with the previous items and their paragraphs in her answer to R2L-*st*.

On the other hand, low scorers surpassed their high-scoring counterparts in their use of strategy (*st*1). Using strategy (*st*1), low scorers resorted to random guessing in order to select answer choices on R2L-*st*.

> *I did not have any choice with this item. I was reading the question* [directions] … and then the answer choices … *while keeping an eye on the time. I selected two answer choices quickly* <*st*1>. (Waleed, RS2I12)

Once Waleed started answering the R2L-*st* item, he noticed the time warning and felt worried about the time. Therefore, he selected two answer choices quickly and with no specific reason.

Discussion.

Research question two (*Are there any differences between high- and low-scorers among subjects in their use of test-taking strategies on the reading tasks and items?*) investigates if there are differences between high- and low-scorers among respondents in their choice and use of test-taking strategies on the TOEFL-iBT reading tasks. The answer to the second research question takes the form of comparisons between high- and low-scoring groups that make use of frequency counts of strategy use. Differences in strategy use between top and bottom scorers were examined across the ten item types.

The high- and low-scoring groups differed in their use of test-taking strategies. High scorers utilized a set of strategies that distinguished them from low scorers across the item types. In general, high scorers benefited from the use of strategies that enabled them to manage their performance well on the reading tasks. Strategy use among high scorers varied depending on the item type; and thus, high scorers did not stick to a certain set of strategies they applied to all item types. This finding points to high scorers' awareness that different item types called for different sets of strategies. For example, Badrah suggested that she usually deals with the BC-*nf* item in a manner different from the other item types:

> *With "except"* [or BC-*nf*] *items, I try to read the paragraph fast first and note mentally these entities that are mentioned, so that when I read the options I can determine which ones are not mentioned in the paragraph* <*nf*4>.

Besides strategy (BC-*nf*4), there are other examples of strategies that point out high scorers' skill of responding to the test items while adopting comprehension and reasoning processes. Evidence that supports this finding comes from the manner in which high scorers responded to item types that required them to read their respective paragraphs. Because the I-*it* items, for example, asked test takers to figure out where a given sentence could fit in a given paragraph in the passage, high scorers chose to read the paragraph and at the same time determine where the sentence could be inserted (I-*it*2). Badrah used strategy (I-*it*2) with Item 11 on Reading Set 2:

> *I chose Option 1 because here it says "Under very cold conditions …," and then "In dry areas, the wind …," so "In dry areas the wind …" is different from "Under very cold conditions …"* <*it*2>. *I felt that this way these three sentences connect with one another.*

Badrah was encouraged to pick the option for the slot just before the sentence starting with *"In dry areas, the wind …"* for two reasons: First, she understood the given sentence

as well as the first three sentences of the paragraph; and second, she noticed that if the given sentence were inserted in this slot, the first four sentences of the paragraph would read coherently.

With the R2L task items (i.e., R2L-*ps* and -*st*) whose aim is to assess test takers' ability to develop an overall understanding of the passage (ETS, 2009), high-scoring test takers used strategies that reflect their attainment of such a high level of understanding. High scorers were able to figure out the intent of the authors of the given passages, recognize rhetorical relationships among different components of the content, and create a mental schema of the major points and ideas presented about the given topics. High scorers' selection of answer choices to R2L-*ps* and -*st* items relied mostly on the understanding they gained as a result of reading or skimming the passage (R2L-*ps*1and R2L-*st*7). Emad expressed this point clearly about his response to R2L-*ps*: *My selection of the choices here was based on what I understood when I skimmed the passage* <*ps*1>, *because I still remember the main ideas*. High scorers were also able to determine which among the answer choices represented important points or ideas in the passage (R2L-*ps*6). Strategy (R2L-*ps*6) was the first attempt that Fahdah took in her response to R2L-*ps*: *I read the answer choices quickly, trying to determine what the main points are* <*ps*6>.

Top-scoring test takers exhibited highly frequent use of confirmation and check strategies. Top scorers exercised more precaution by confirming that they had picked the right answer, even when they were positive about the correctness of their answers. About her answer to BC-*pr*, Fahdah said: *I read the sentences before and after the sentence that has "them"* [the pronoun] <*pr*2>. *I was already convinced from my first reading of the*

paragraph that the answer is "masses of ice". Fahdah's use of confirmation strategy points to a common characteristic among top scorers, namely the use of means that relied on skills of reading comprehension. Instances of confirmation strategy among top scorers, including rereading the whole paragraph or a portion of it (BC-*fi*, I-*it*, and I-*rp*), reading or rereading the sentence with the target word (BC-*v*) or pronoun (BC-*pr*), rereading the highlighted sentence (BC-*ss*), or referring to the passage (R2L-*st*) were the first choices.

High scorers were also observed to use check strategies when trying to determine which answer choices to select in response to R2L-*st*. Thus, they checked some or all of the answer choices on R2L-*st* by referring to the passage in order to find out if a given answer choice represented a "constructive" or "destructive" process (R2L-*st*2). Deeb benefited from strategy (R2L-*st*2):

> *I was not sure about Choice 4, so I checked the passage <st2>and found it to be "constructive" from "In contrast, the roots of grasses … to hold loose soil fragments together, thereby helping to prevent erosion by the wind."*

Top scorers' use of confirmation and check strategies indicates their propensity to evaluate and regulate their strategy use on the reading tasks. This finding converges with those of previous research (N. Anderson, 1991; Nikolov, 2006; Phakiti, 2003; Tian, 2000) in regard to the superior levels of strategic awareness and monitoring that high scorers tend to exhibit on tests of reading comprehension.

Top scorers treated background knowledge as a resource to draw on especially with BC-*v*. When working with BC-*v*, resorting to background knowledge served either one of two main goals for top scorers, that is, to think of or to retrieve the meaning of the target word or choose a preliminary option. The reason why top scorers employed such a

strategy mostly with the BC-*v* items was that in contrast to other item types, BC-*v* items were not totally dependent on the content of the passage. Although with low frequencies, top scorers reported the use of background knowledge with other item types (incl., BC-*fi*, BC-*ss*, and I-*rp*) especially when they wanted to relate the content of the item and/or the options to what they already knew. A possible explanation is that top scorers had a high level of understanding of what they read in the text, which allowed them to relate what they understood to their background knowledge. In this regard, Emad reported: *I read the paragraph, and because I am already familiar with the US political system, the Democrats do not want a state religion* <BC-*fi*8>.

On the other hand, test takers in the low-scoring group used strategies in a manner that spoke of their low levels of performance and scoring. Low-scoring test takers' strategies were almost the same across different item types; and thus, low scorers adhered to a certain set of strategies regardless of the item type. This suggests that low scorers had either limited awareness of strategies that are applicable to various item types or deficient ability to use such strategies. Among the most frequent strategies that low scorers employed with several item types were the uses of clues and guessing to choose an option. Low scorers resorted to these strategies mainly when unable to use their reading skills to figure out the right answer on such item types as BC-*v*, I-*bi*, and I-*it*. These item types were especially challenging for low scorers. Saja described her encounter with I-*bi*: *Inference items are never easy, and my time was getting closer to the end, so I chose the answer out of my free choice.*

Low-scoring test takers also made use of clues and guessing when they were in a state of hesitation about the answer and needed to decide on an option, as was evident

with BC-*ss* and I-*it*. Bottom scorers suggested that the main source of complexity with BC-*ss* was the similarity of the options (c.f., C. Taylor et al., 1998). Such a factor caused most bottom scorers to look for clues to determine which option could replace the highlighted sentence. In his answer to the BC-*ss* item on Reading Set 2, Ziyad used the fact that one of the options shared the mention of a definite time with the highlighted sentence as a clue that helped him decide on the option. When answering the I-*it* item on Reading Set 2, Tariq mentioned that he decided on Option 3 because the choice of Option 4 would make the given sentence separate two consecutive sentences, the first of which had "landscapes" which had to do with "tree roots" in the second sentence.

Most low scorers selected answer choices on the R2L items using random guessing. Low scorers were motivated to make use of random guessing with the R2L items for the reason that by the time they reached these items, they had consumed most of the allotted time. About his response to R2L-*ps*, Yasser suggested that he selected answer choices in order (i.e., 2, 3, and 4) after skipping the first one, because when he got to this item, he had three seconds left of the allotted time. Tariq suggested that he guessed his selections of answer choices on R2L-*st* because he did not know how "constructive" and "destructive" processes differ from each other. Similarly, findings of previous studies (e.g., N. Anderson et al., 1991; Nevo, 1989; Tian, 2000; Upton & Lee-Thompson, 2001; Yamashita, 2003) have indicated that test takers with low levels of proficiency and/or test performance utilized clues and guessing on assigned reading tasks.

Poor reading skills of low scorers encouraged the use of other strategies, some of which can be described as test-wiseness in nature. These strategies included rereading the given paragraph or a portion of it (BC-*fi*3, I-*bi*3, and I-*rp*3), or the highlighted sentence

(BC-*ss*6), and matching the content of the given paragraph or sentence, or elements of it (e.g., a key word/phrase), and the options (BC-*fi*7, BC-*ss*3, I-*bi*7, and I-*rp*8). Low scorers' frequent use of strategies involving rereading and matching with BC-*fi*, BC-*ss*, I-*bi*, and I-*rp* item types suggests that these item types posed difficulty for this scoring group, especially I-*bi* which scored too low on the item difficulty index. One motive to use strategies of rereading and reading more than once was because of unfamiliar vocabulary in the text, as Waleed mentioned about his response to the BC-*ss* item on Reading Set 2: *I read the highlighted sentence more than once, trying to understand it. I understood it though it has some unknown vocabulary <ss*6>, *for example, "epitome of permanence"*. Tariq noted that the similarity of the options on the BC-*ss* item on Reading Set 2 caused him to read the highlighted sentence more than one time: *This item requires you to find the closest statement in the options to the highlighted sentence, so the options are always close to one another … I read the highlighted sentence more than one time <ss*6>.

> Saja reported using the matching strategy with the BC-*fi* item on Reading Set 2:
>
> *I read the paragraph. Then, I read the options while trying to find out a key word for the question* [item]. *I read in the paragraph "… crashing into each other …" and I used this as key information to choose Option 3 <fi*7>. *Also, this sentence* [the source of critical information] *and Option 3 have the same meaning, because "crash" is like "hit" <fi*7>.

Low scorers employed another strategy, namely attempting to locate the critical information using a key word or phrase (I-*bi*6), for the reason of item difficulty. The key word or phrase in the description of this strategy represents an element that is shared by the critical source of information in the paragraph and an option. Yasser described a desperate use of strategy (I-*bi*6) in his answer to the I-*bi* item on Reading Set 2:

> *I saw that the author focuses on "height" so much … I read the options, trying to fix them well in my mind. I read the paragraph, trying to locate information on "height" and link it to an option* <*bi*6>.

Bottom scorers were inclined to read the surrounding context besides reading the sentence that had either the target word on the BC-*v* items or the highlighted phrase on the I-*rp* items. For bottom scorers, reading the sentence containing the vocabulary item or the highlighted phrase by itself did not suffice to figure out the meaning of the vocabulary item or the reason why the author mentioned the phrase highlighted in the sentence. Most bottom scorers attempted to read the sentence with the target word or the highlighted phrase first, but then they had to read a wider context by reading the surrounding sentences. Saja reported that she followed the same strategic moves in her answer to the RS1BC-*v* item on Reading Set 1, and so did Waleed about his answer to the I-*rp* item on Reading Set 2.

Another strategy that bottom scorers were motivated to use with BC-*v* was replacing the target word with each word among the options and judging the suitability of the replacement in the context of the target word (BC-*v*5). Ziyad used this strategy with the BC-*v* item on Reading Set 2: *The meanings of all four options were not clear, so I used the way each option sounds in the context of the* [target] *word to decide if it is right or wrong.* Bottom scorers followed a shortcut move when responding to BC-*nf* by attempting to locate each option in the paragraph and at the same time eliminating the options they found (*nf*1). It was because of the simplicity of this strategy that Yasser reported: *This was an easy item. Three of these options are mentioned in the paragraph and one, for certain, is not mentioned* < *nf*1>. Other researchers reported that their respondents, who were engaged in response to reading tasks, exhibited frequent use of

such strategies as rereading of textual information (Cohen & Upton, 2006), matching elements of the item content (Rupp et al., 2006), and word substitution with vocabulary items (Upton & Lee-Thompson, 2001).

Although both high- and low-scoring groups used certain strategies almost equally, they differed in their intentions behind the use of such strategies. High scorers reported that they eliminated certain options (*ss*4) with the BC-*ss* items because these items were challenging, as Fahdah said about the BC-*ss* item on Reading Set 1: *This is one of the items that gave me hard time*. High scorers reported using the elimination strategy frequently with R2L-*ps*, since test takers need to differentiate between important ideas and less important or irrelevant ones (ETS, 2009). Ahmad, for example, suggested that he eliminated certain answer choices on R2L-*ps* (*ps*3) on the basis of his understanding of the passage. In contrast to high scorers, low scorers reported that they used the elimination strategy with options and answer choices when they did not know what they meant. For this reason, Tariq dropped Option 1 from his consideration when answering the BC-*v* item on Reading Set 2: *I thought that Option 2 "regions" would more likely be the answer than Option 3 "remains." I did not know the meaning of Option 1 "resemblances".*

Another strategy the two scoring groups shared, but with different goals, was the strategy of reading the options first. High scorers started with the options on I-*rp* in order to be familiar with the options or choose a preliminary option before they read the sentence with the highlighted phrase. Deeb suggested that he started with the options on the I-*rp* item on Reading Set 1 so as to see if he could find a suitable option to choose. On the contrary, low scorers read the options first on BC-*nf* because they wanted to scan the

paragraph for the options in an attempt to identify the exception (*nf*5), as Ziyad did in his answer to BC-*nf* on Reading Set 1: *I read the options <nf5>. I scanned the paragraph for the options one by one, and eliminated some options to determine the answer*.

Both high- and low-scoring groups reported using clues in a manner that set them apart, specifically with two distinct item types. High scorers used clues with BC-*ss* to choose an option because this item type was the most challenging for high scorers. Seven out of ten reports by the high-scoring group suggested using clues in response to BC-*ss*. Deeb described a problem he encountered with the BC-*ss*: *When the author produces paraphrases* [options], *he uses synonyms and changes the voice*, which made him choose an option using "although" which shared the same semantic relation with "but" in the highlighted sentence (*ss*2)—both "but" and "although" conveyed concession. Low scorers used clues copiously as mentioned earlier, and the reason for such extensive use becomes even clearer with BC-*pr*. Waleed had difficulty with BC-*pr* because most of the words in the sentence with the pronoun were not familiar to him, and so he felt urged to use a clue: *I read the sentence from the beginning, but stopped because there are vocabs unknown to me… I used the comma before "carrying with them" because it indicates reference to the last word (i.e., valleys) <pr*5>. Test takers in Tian's (2000) study, who responded to the reading subtest of the TOEFL-PBT, demonstrated use of the same three strategies shared by high- and low-scoring groups in this study, regardless of their levels of performance and scoring.

Research question three.

What aspects of effective test-taking strategy use do subjects tend to employ with the reading tasks and items? In order to answer research question three, strategy clusters

that were associated with the selection of the correct answers to item types were examined for aspects that would reflect efficacy of strategy use. A strategy cluster represents a group of strategies that are used in combination. As such, a strategy cluster can be linked to successful versus unsuccessful strategy use in language learning and testing (c.f., Cohen, 2005; N. Anderson, 2005; Oxford, 1996). Presumably, successful performance on each test item entails effective strategy clustering which intrinsically involves logical sequencing of strategies within each strategy cluster. It is this sense of having logically sequenced strategies that determines the extent to which strategy use can be effective, as suggested by experts in learner strategies (Cohen, 2005).

The analysis for the third research question focused on patterns of strategy sequences across all item types. Since there were items with single appearance on the reading sets (incl., BC-*nf*, BC-*pr*, R2L-*ps* and R2L-*st*), strategy sequences used by individual test takers with these item types were considered in the analysis. Strategy clusters in this study were found to comprise strings or sequences of two to three strategies. Strategies in sequences cluster around one to three strategies. Strategy clusters that test takers employed with each item type are graphed using flowcharts and discussed here, and verbal data are used to illustrate beneficial aspects of strategy sequences. Flowcharts were produced using yEd, designed by yWorks GmbH (2010). Strategy descriptions are included in the discussion.

Strategy abbreviation(s) specific to this section:

- (RS#TT-*i*I#): Serial number of reading set (1 for Reading Set 1, and 2 for Reading Set 2), task type, item type and number, respectively. (E.g., RS1BC-*fi*I7 denotes

Item 7 as a Basic Comprehension-*factual information* or BC-*fi* item on Reading Set 1).

- (TT-*i*#→ *i*#): Sequence of two strategies with one item type. (E.g., I-*bi*1→*bi*2 denotes a sequence of strategies (*bi*1 and *bi*2) among strategies used with Inferencing-*basic inference* or I-*bi* items).

Effective strategy clusters by task and item type.

Basic comprehension-vocabulary (BC-v). As shown in Figure 13 below, test takers made use of four strategy sequences with the two BC-*v* items.

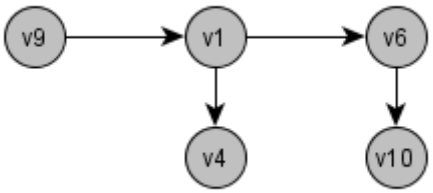

Figure 13. Strategy cluster with BC-*v* items. (*Note*. arrow = 2 strategies in 1 sequence).

Strategy sequences associated with test takers' responses to BC-*v* were as follows.

1. (*v*1→*v*6). Test takers making use of this sequence read the sentence containing the target word to figure out what the word meant in the context of the sentence (*v*1). Then, test takers eliminated certain options as incorrect (*v*6). For example, Participant (23) followed this sequence in his answer to RS1BC-*v*I1:

> *I read the sentence that has the word <v1>. Then, I read the options. Option 1 "frequently" is not the right meaning because it implies something occurs at intervals <v6>. For Option 3 "rapidly," there is no connection between something fast and something that cannot be measured <v6>. For Option 4, I did not expect it to be right <v6>. I thought that "immeasurably" means 'something so huge.'*

2. (*v*9→*v*1). This sequence typified response behaviors of test takers who first consulted their background knowledge to think of or retrieve an exact or close meaning of the

target word (*v*9). And then, they read the sentence with the target word in order to confirm the word meaning they knew (*v*1). About his answer to RS2BC-*v*I4, Participant (17) mentioned: *I knew "relics" from my background knowledge to mean 'history'* <*v*9>. *I read the sentence with the* [target] *word and understood from "eroded relics of much higher mountain …" that "relics" means Option 3 "remains"* <*v*1>.

3. (*v*1→*v*4). This sequence was typical of test takers who could not figure out the word meaning by just reading the sentence containing the target word (*v*1). Therefore, they had to read the sentence with the target word along with the surrounding context (*v*4). The surrounding context refers to the sentence before the sentence with the target word, or the one after it, or both. For example, participant (19) said about RS2BC-*v*I4 that: *It was a difficult item. I could not understand the sentence* [with the target word] <*v*1>, *so I read the sentences before and after it* <*v*4>.
4. (*v*6→*v*10). When test takers used this sequence, they first eliminated those options that sounded incorrect or irrelevant (*v*6) based on the contextual meaning of the target word. Next, when test takers were left with two or more options, they decided on one option using semantic clues (*v*10). Participant (24) followed this sequence with RS2BC-*v*I4:

> *I read the sentence that has the word and the sentences before and after it to figure out the meaning* [of the target word] … *I eliminated Options 1 and 4 because they seemed to me irrelevant to the question* <*v*6>. *I hesitated between Options 2 and 3, but chose Option 3 using "eroded" as a clue linked more to "remains"* [Option 3] *than "regions"* [Option 2] <*v*10>.

Basic comprehension-factual information (BC-fi). As shown in Figure 14 below, test takers made use of four strategy sequences with the two BC-*fi* items.

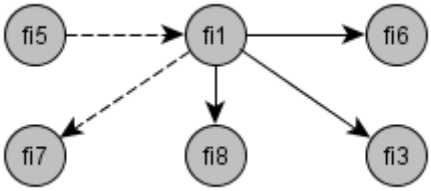

Figure 14. Strategy cluster with BC-*fi* items. (*Note*. solid arrow = 2 strategies in 1 sequence, and dashed arrows = 3 strategies in 1 sequence).

Strategy sequences associated with test takers' responses to BC-*fi* were as follows.

1. (*fi*1→*fi*6). Test takers using this sequence first read the paragraph referred to by the BC-*fi* item (*fi*1). Then, test takers eliminated those options they found to be either incorrect or irrelevant (*fi*6) based on the information stated in the paragraph. For example, Participant (7) reported that his answer to RS1BC-*fi*I7 followed this sequence:

   > *First, I read the paragraph to find the answer <fi*1>. *Then, I eliminated Option 1 because it is not related to the information content of this item <fi*6>. *I also eliminated Option 2 'because they try to limit the role of the government' and also Option 4 'because they are trying to use people to achieve their own profit' <fi*6>.

2. (*fi*1→*fi*3). This sequence involved test takers' reading of the paragraph referred to by the BC-*fi* item (*fi*1), and then, rereading of the paragraph or a portion of it so as to locate the source of critical information (*fi*3). About his answer to RS2BC-*fi*I5, Participant (19) reported: *I read the paragraph and stopped at "Some mountains are formed as a result of …" <fi*1>. *I read this* [portion of the paragraph] *more than one time <fi*3>.

3. (*fi*1→*fi*8). When test takers used this sequence, they basically read the given paragraph (*fi*1). And then, they consulted background knowledge in relation to the information

presented in the paragraph (*fi*8). In his report about how he answered RS2BC-*fi*I5, participant (1) said: *I read the* [given] *paragraph <fi*1*>. I used my understanding of the paragraph and my knowledge in geography 'like when plates collide,' mountains are formed <fi*8*> and in the paragraph it says "plates crashing into each other."*

4. (*fi*5→*fi*1→*fi*7). Test takers who used this sequence read the options first (*fi*5). Then, they read or skimmed the given paragraph (*fi*1) while using the content of the options to locate the source of the critical information. After that, test takers matched key words or phrases in the options and the source of critical information in the paragraph to determine the required information (*fi*7). Participant (2) followed this approach with RS1BC-*fi*I7, as he reported:

> *First, I read the question and skimmed the options <fi*5*>. I read the part of the paragraph that is related to this item <fi*1*>, and I felt that the answer is given directly "destroying monopolies" <fi*7*>, so I chose Option C "destroying monopolies."*

Participant (2) in this instance read the upper half of the given paragraph while he was answering the item before; consequently, with the BC-*fi* item he chose to read the bottom half of the paragraph.

Basic comprehension-negative fact (BC-nf). As shown in Figure 15 below, test takers made use of two strategy sequences with the BC-*nf* item.

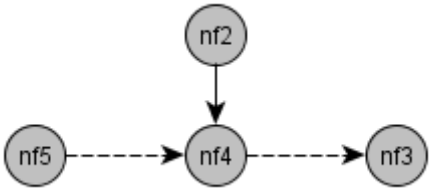

Figure 15. Strategy cluster with BC-*nf* item. (*Note*. solid arrow = 2 strategies in 1 sequence, and dashed arrows = 3 strategies in 1 sequence).

Strategy sequences associated with test takers' responses to BC-*nf* were as follows.

1. (*nf*2→*nf*4). Test takers using this sequence first read the paragraph referred to by the BC-*nf* item in order to identify the exception (*nf*2) through the elimination of options as either incorrect or mentioned in the paragraph (*nf*4). For example, Participant (22) said about his answer to RS1BC-*nf*I10:

   > *I read the paragraph, and read that "the Democrats… among farmers… and workers … planters…." <nf2>. Option 1 is not right <nf4>. For Option 2, … "rising entrepreneurs" is mentioned* [in the paragraph] *<nf4>. Option 4 is also mentioned <nf4>.*

2. (*nf*5→*nf*4→*nf*3). This sequence was used by test takers who chose to read the options first to have familiarity with them (*nf*5). Then, test takers read or skimmed the paragraph and at the same time eliminated certain options because such options were either incorrect or mentioned in the paragraph (*nf*4). Participant (7) confirmed his answer by rereading the relevant portion of the paragraph (*nf*3). Participant (7) followed this sequence with RS1BC-*nf*I10:

   > *I skimmed the options <nf5>, and I read the paragraph carefully … eliminated some of the options like "workers unhappy with …"* [Option 1] *because there is no need for it <nf4>, I considered "rising entrepreneurs"* [Option 3] *possible … I chose Option 4 'because Democrats tried to help individuals participate in the market, while the Whigs helped those people who already had the market.' I reread the bottom half of the paragraph <nf3>, and changed my answer to Option 2.*

Basic comprehension-pronoun reference (BC-pr). As shown in Figure 16 below, test takers made use of three strategy sequences with the BC-*pr* item.

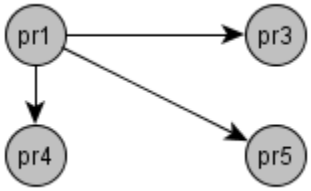

Figure 16. Strategy cluster with BC-*pr* item. (*Note*. arrow = 2 strategies in 1 sequence).

Strategy sequences associated with test takers' responses to BC-*pr* were as follows.

1. (*pr*1→*pr*3). This sequence was followed by test takers who first read the sentence containing the target pronoun (*pr*1). And then, they eliminated options that sounded to them incorrect based on the information content of the sentence (*pr*3). For example, about his answer to RS2BC-*pr*I8, Participant (18) reported: *I read the sentence that has the pronoun <pr*1>. *I found that the author talks about "glaciers" in "cold areas," so I eliminated Option 1 "cold areas" <pr*3>.
2. (*pr*1→*pr*5). When test takers used this sequence, they first read the sentence containing the pronoun (*pr*1), and then used clues to choose an option (*pr*5). Participant (25) followed this sequence with RS2BC-*pr*I8: *I read the sentence* [containing the target pronoun] <*pr*1>. *I suspected the answer to be "rock debris," but the sentence ends in a period, so the pronoun must have to do with the sentence* [clause] *before* [Option 2] <*pr*5>.
3. (*pr*1→*pr*4). Test takers using this sequence first read the sentence with the target pronoun (*pr*1) more than once to identify the referent of the pronoun (*pr*4). About his answer to RS2BC-*pr*I8, Participant (19) said:

> *I read the sentence* [with the target pronoun] <*pr*1> *and the one before till I got to the word "debris." I read it* [the sentence with target pronoun] *the first time and found the answer to be Option 3 "valleys," I read it a second time*

and found the answer to be Option 2 "masses of ice" <pr4>, I read it a third time and found the answer to be Option 4 "rock debris" <pr4>… I chose Option 2 "masses of ice."

Basic comprehension-sentence simplification (BC-ss). As shown in Figure 17 below, test takers made use of four strategy sequences with the BC-*ss* items.

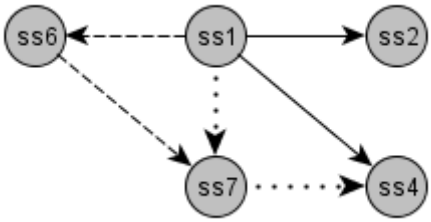

Figure 17. Strategy cluster with BC-*ss* items. (*Note*. solid arrow = 2 strategies in 1 sequence, dashed arrows = 3 strategies in 1 sequence, dotted arrows = 3 strategies in 1 sequence).

Strategy sequences associated with test takers' responses to BC-*ss* were as follows.

1. (*ss*1→*ss*2). Test takers who used this sequence read the highlighted sentence first (*ss*1) and then used clues to choose an option (*ss*2). Participant (5) reported that he followed this sequence in his answer to RS2BC-*ss*I9:

   I read the highlighted sentence <ss1>and it was easy. Then, I read the options; they were close. I thought about Option 2 which contains "although" because of "but" in the highlighted sentence; "although" is a synonym of "but" … Therefore, I chose Option 2 <ss2>.

2. (*ss*1→*ss*4). This sequence was used by test takers who read the highlighted sentence first (*ss*1), and then used their understanding to eliminate certain options, considering these options either incorrect or irrelevant in the light of the information content of the highlighted sentence (*ss*4). This was the case with Participant (22) when answering RS2BC-*ss*I9:

> *When I read the highlighted sentence <ss1>, I encountered "epitome." I do not know what this word means; I guessed its meaning to be like 'an example of permanent things' that do not change over time. The paragraph does not point to a certain stage in the mountain's lifetime … as in Option 1 <ss4>.*

3. (*ss*1→*ss*6→*ss*7). Test takers using this sequence read the highlighted sentence (*ss*1) more than once (*ss*6) in order to grasp the general meaning of the highlighted sentence (*ss7*). About his answer to RS2BC-*ss*I9, Participant (10) reported:

> *I read the highlighted sentence <ss1>more than once <ss6>…. I understood from the highlighted sentence that 'mountains may be affected by natural causes although they appear huge, but they age for a short period of time in geological terms' <ss7>. My answer was very close to this meaning, which is Option 2.*

4. (*ss*1→*ss*7→*ss*4). This sequence typified response behaviors of test takers who read the highlighted sentence first (*ss*1) with the intention of understanding its general meaning (*ss7*). Then, test takers used their grasp of the general meaning of the highlighted sentence to eliminate options they found incorrect (*ss*4). The sequence of the three strategies was followed by Participant (13) in her response to RS1BC-*ss*I11:

> *The highlighted sentence describes the case of both 'the Whigs and the Democrats' <ss1 & ss7>. When I read the options … I eliminated Option 3 initially because it reverses the structure of the* [highlighted] *sentence <ss4>, which meant to me that the meaning is changed. I chose Option 1 first, but changed to Option 4 because Option 1 does not express the whole meaning or the right meaning of the highlighted sentence.*

Inferencing-basic inference (I-bi). As shown in Figure 18 below, test takers made use of three strategy sequences with the I-*bi* items.

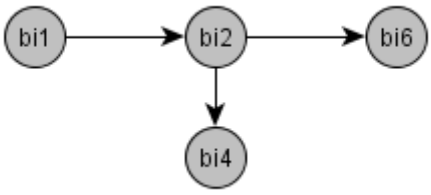

Figure 18. Strategy cluster with I-*bi* items. (*Note*. arrow = 2 strategies in 1 sequence).

Strategy sequences associated with test takers' responses to I-*bi* were as follows.

1. (*bi*1→*bi*2). This sequence was followed by test takers who read the options first (*bi*1) so that they could locate the critical information when they read the paragraph referred to by the I-*bi* item (*bi*2). Participant (17) reported that he made use of this sequence in his response to RS2I-*bi*I3:

   > *I read the options first <bi*1*>. I skimmed the paragraph here* [pointing to the paragraph on the screen]. *While skimming the paragraph, I had in mind the figures and names of mountains that I read in the options. I used these as key words or information to find where the answer is in the paragraph <bi*2*>.*

2. (*bi*2→*bi*4). This sequence involved test takers' reading of the paragraph referred to by the I-*bi* item first (*bi*2), and then, elimination of options that test takers considered either incorrect or irrelevant (*bi*4) in view of the information content in the paragraph. For example, Participant (22) said about his answer to RS1I-*bi*I9:

   > *I read the paragraph, and read that "it should be used to protect individual rights and public liberty …" <bi*2*>. I deduced that the Whig party was not focusing on "issues of public liberty" only (or Option 1) <bi*4*>. Option 2 is irrelevant <bi*4*>. Option 3 is also irrelevant <bi*4*>.*

3. (*bi*2→*bi*6). Test takers who employed this sequence read the given paragraph (*bi*2) while attempting to locate the source of critical information by means of a key word or phrase (*bi*6). Participant (24) followed the (*bi*2→*bi*6) sequence while responding to RS2I-*bi*I3: *I read the paragraph <bi*2*> and at the same time tried to find the answer using "the mountains of the Himalayas"* [in the item stem or question] <*bi*6>.

Inferencing-insert text (I-it). As shown in Figure 19 below, test takers made use of five strategy sequences with the I-*it* items.

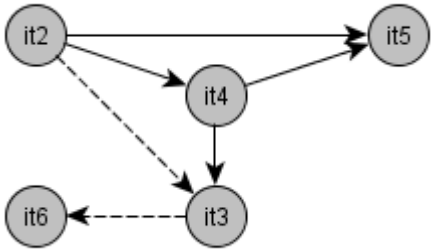

Figure 19. Strategy cluster with I-*it* items. (*Note*. solid arrow = 2 strategies in 1 sequence, and dashed arrows = 3 strategies in 1 sequence).

Strategy sequences associated with test takers' responses to I-*it* were as follows.

1. (*it*2→*it*4). This sequence was used by test takers who first read the paragraph referred to by the I-*it* item (*it*2) and at the same time tried the given sentence in one or more of the slots in the paragraph to determine where it best fitted (*it*4). Participant (12) made use of this sequence in her response to RS1I-*it*I12: *I read the paragraph <it*2*>. I noticed that the given sentence tells about something new, and so I tried it in the first position* [Slot 1] *<it*4*> and checked its suitability in this position.*
2. (*it*2→*it*5). Test takers using this sequence read the paragraph referred to by the I-*it* item (*it*2) and at the same time eliminated the options for the slots where the context sounded irrelevant to the theme of the given sentence (*it*5). The (*it*2→*it*5) sequence was followed by Participant (1) in his response to RS2I-*it*I11:

   > *The second sentence* [in the paragraph] *talks about dry areas <it*2*>, and the given sentence should be placed before a sentence which talks about different climate, may be humid or dry. The next sentence* [after the given sentence] *would be an example like "In dry areas." For Options 3 and 4, they do not fit because the sentence before* [Slots 3 & 4] *talks about 'living things' like "roots," which are totally different from the discussion of climate <it*5*>.*

3. (*it*4→*it*3). Test takers who used this sequence tried the given sentence in one or more of the slots in the given paragraph (*it*4). At the same time, test takers tried to identify a

logical connection between the given sentence and the sentence(s) before and/or after the given slot (*it*3). About his answer to RS1I-*it*I12, Participant (3) said:

> *I tried* [the given sentence] *in each potential place* [slot] <*it*4>. *I chose Option 1* [Slot 1] *because the previous sentence introduces something new, I mean "the Whig party," and the next* [given] *sentence discusses the "Whig party"* <*it*3>.

4. (*it*4→*it*5). When test takers used this sequence, they first tried the given sentence in one or more of the slots in the given paragraph (*it*4). At the same time, test takers eliminated the options for the slots where the context seemed distant from the context of the given sentence or irrelevant to its theme (*it*5). For example, Participant (17) said that in his answer to RS1I-*it*I12:

> *I tried to insert the given sentence into each potential place* [slot] *in the paragraph* <*it*4>. *I felt that Option 1 made sense to me, Option 2 did not fit* <*it*5>*, and Option 3 sounded to me like the* [given] *sentence is placed further in the text than it should be* <*it*5>.

5. (*it*2→*it*3→*it*6). This sequence was followed by test takers who employed the strategy of confirming the selected answer. First, test takers read the paragraph referred to by the I-*it* item (*it*2). Then, they tried to find a logical connection between the given sentence and the sentence(s) before and/or after a given slot to choose an option (*it*3). After that, they confirmed their answer by checking the other options or slots to ensure that none of them accepts the given sentence (*it*6). When reporting her answer to RS1I-*it*12, Participant (13) said:

> *I found that the given sentence fits in Place 1* [Slot 1] *because the first sentence says "… to form the Whig party"* <*it*2> *and the given sentence begins with "This party," so I chose Option 1* [Slot 1] <*it*3>. *I checked the other options* <*it*6>*, but felt that Option 1 is the best.*

Inferencing-rhetorical purpose (I-rp). As shown in Figure 20 below, test takers made use of three strategy sequences with the I-*rp* items.

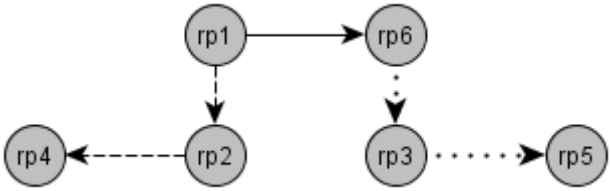

Figure 20. Strategy cluster with I-*rp* items. (*Note*. solid arrow = 2 strategies in 1 sequence, dashed arrows = 3 strategies in 1 sequence, dotted arrows = 3 strategies in 1 sequence).

Strategy sequences associated with test takers' responses to I-*rp* were as follows.

1. (*rp*1→*rp*6). Test takers using this sequence read the options first in order to have familiarity with them (*rp*1) before reading the paragraph referred to by the I-*rp* item (*rp*6) to figure out the required information. For example, Participant (3) reported about his answer to RS2I-*rp*I6: *I read the options <rp*1*>, but the options are a bit similar. Then, I read the paragraph for more specific information <rp*6*>, and I was able to deduce the meaning.*
2. (*rp*1→*rp*2→*rp*4). When test takers used this sequence, they first read the options to familiarize themselves with them (*rp*1). Then, test takers read the sentence containing the highlighted phrase and the surrounding context to figure out the required information (*rp*2). After that, they confirmed their answer by rereading the paragraph or the relevant portion of it (*rp*4). For example, Participant (24) said about his answer to RS2I-*rp*I6:

> *I read the options while contrasting them <rp*1*>. I read the sentence that has the highlighted phrase, and the sentence before, and the one after <rp*2*>… it mentions that "carbon dioxide reacts with the rain and forms a weak acid … this chemical attacks rocks." And, I chose the answer directly. I chose Option 1. I continued reading to check because there might be some other information which is the actual answer <rp*4*>.*

3. (*rp*6→*rp*3→*rp*5). This sequence was followed by test takers who read the paragraph referred to by the I-*rp* item (*rp*6). Then, test takers read the relevant portion of the paragraph more than once (*rp*3) in order to figure out the required information. After that, test takers eliminated those options they found incorrect (*rp*5) based on what they understood. Participant (19) followed this approach with RS2I-*rp*I6: *The paragraph is short, so I had my time reading it* <*rp*6>. *I had to reread the sentence before the sentence containing the highlighted phrase* <*rp*3>*… the author does not intend to give an example* [Option 3] <*rp*5>.

Reading to learn-prose summary (R2L-ps). As shown in Figure 21 below, test takers made use of three strategy sequences with the R2L-*ps* item.

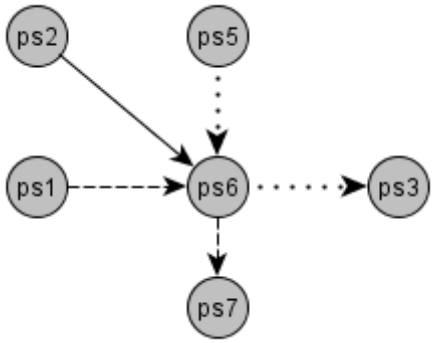

Figure 21. Strategy cluster with R2L-*ps* item. (*Note*. solid arrow = 2 strategies in 1 sequence, dashed arrows = 3 strategies in 1 sequence, dotted arrows = 3 strategies in 1 sequence).

Strategy sequences associated with test takers' responses to R2L-*ps* were as follows.

1. (*ps*2→*ps*6). Test takers who used this sequence relied on the understanding they had developed as a result of going through the previous items and reading their respective paragraphs (*ps*2) in their selection of answer choices that represented important or summary points of the passage (*ps*6). For example, Participant (1) reported that he

followed the (*ps*2→*ps*6) sequence in his answer to RS1R2L-*ps*I13: *I read the answer choices, and tried to remember if any of the ideas I understood from answering previous items <ps*2*> are repeated here <ps*6*>.*

2. (*ps*1→*ps*6→*ps*7). Other test takers used their understanding from skimming or reading the passage (*ps*1) at the outset of their dealing with Reading Set 1 to select answer choices that represented important or summary points of the passage (*ps*6). After following the (*ps*1→*ps*6) sequence, Participant (22) confirmed his selections by checking the unselected answer choices to see if any of them was better than any of the selected choices (*ps*7). About his response to RS1R2L-*ps*I13, Participant (22) said:

> *My selection of the choices here is based on my understanding from my skimming of the passage <ps*1*> because I still remember the main ideas <ps*6*> … I thought, for Option 1, that the two parties were formed because of the changes that existed at the time <ps*6*>. After I selected the other options, I remembered from my reading of the passage that other options could be among the main ideas of the passage <ps*7*>… At last, I decided about Choices 1, 5, and 6 because they represent important points.*

3. (*ps*5→*ps*6→*ps*3). This sequence was used by test takers who used the introductory sentence (*ps*5) to guide their selection of answer choices that represented important or summary points of the passage (*ps*6). Then, test takers eliminated those answer choices that they considered inappropriate to be summary points (*ps*3). Participant (10) reported that he followed this sequence in his answer to RS1R2L-*ps*I13:

> *I started with the sentence that represents the 'topic sentence'* [introductory sentence] *in the paragraph <ps*5*>, so I started with 'the President' in Choice 2 "During Andrew Jackson's two terms …." I selected this choice because it has the main meaning of the passage <ps*6*> … I eliminated "The Democratic party primarily represented …"* [Choice 3] *because it does not represent a main point in the passage <ps*3*>.*

Reading to learn-schematic table (R2L-st). As shown in Figure 22 below, test takers made use of three strategy sequences with the R2L-*st* item.

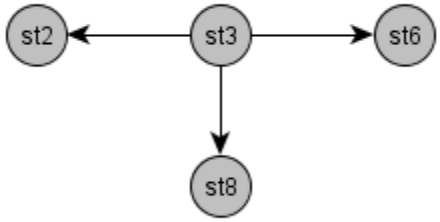

Figure 22. Strategy cluster with R2L-*st* item. (*Note*. arrow = 2 strategies in 1 sequence).

Strategy sequences associated with test takers' responses to R2L-*st* were as follows.

1. (*st*3→*st*2). Test takers making use of this sequence used the understanding they had developed as a result of going through the previous items and reading their respective paragraphs (*st*3) and supplemented it by referring to the passage to determine whether a given answer choice represented a "constructive" or "destructive" process (*st*2). Participant (24) followed this sequence in his answer to RS2R2L-*st*I12:

   > *I felt hesitant about Choice 3, and the first three choices in general, because "Wind-driven sand" can be "constructive" and "destructive" based on a previous item which mentions this <st3>. I read all answer choices while trying to keep them in mind. I referred to the passage to see which choices are "constructive" and which are "destructive" <st2>.*

2. (*st*3→ *st*6). This sequence was followed by test takers who used the understanding they had developed as a result of going through the previous items and reading their respective paragraphs (*st*3) to eliminate answer choices they thought were either unmentioned or unimportant in the passage (*st*6). For example, about his response to RS2R2L-*st*I12, Participant (11) mentioned:

   > *I used my understanding of the paragraphs I read in the previous items to answer this item <st3> … First, I started with "destructive processes" which are three … Choice 2 is not mentioned <st6> … Choice 5 is not mentioned by itself as a factor* [process]*; there is something else besides it, so I eliminated it <st6>.*

3. (*st*3→*st*8). This sequence was followed by test takers who used the understanding they had developed as a result of going through the previous items and reading their respective paragraphs (*st*3) to select answer choices that sounded certain or direct (*st*8). Participant (23) followed this sequence in his response to RS2R2L-*st*I12:

> *I went through the previous items, which increased my retention of information I read in the passage <st3>. I remembered that there are three causes of mountain formation: 'The movement of the Earth's crustal plates,'… "Earthquakes" is mentioned in a direct way <st8>; it is listed as one of the reasons for mountain formation … "Wind-driven sand" is sure <st8>.*

Discussion.

Research question three (*What aspects of effective test-taking strategy use do subjects tend to employ with the reading tasks and items?*) explores aspects of strategy use which aided test takers' performance on the TOEFL-iBT reading tasks and items. The answer to the third research question was attained through examination of strategy clusters that seemed to have led test takers to make the right choices in their answers to the test items. The findings of this study confirm previous theory and research that have pointed out that strategy clusters determine the extent to which strategy use is effective (e.g., Cohen, 2005; N. Anderson, 2005; Oxford, 1996), which in turn accounts for high versus low test performance and scores in the field of language testing and assessment (e.g., N. Anderson, 1991; Nikolov, 2006; Yien, 2001; Yoshizawa, 2002).

In general, strategy clusters that test takers made use of in this study had as their nuclei (or source strategies in strategy clusters) strategies that were used highly frequently among test takers, according to strategy coding schemes. In keeping with the answer to the first research question, source strategies in strategy clusters were by and large compatible with the various item types on the reading tasks. Source strategies also

possessed a high level of flexibility in strategy sequencing in that they almost accepted any other strategy in a sequence without a mediating strategy in between. Most test takers were able to combine with source strategies other strategies (or attached strategies) to form sequences of strategies that were logically connected.

For example, test takers who read the paragraph or a relevant portion of it with BC-*fi,* or strategy (BC-*fi*1), for example, attached to it four strategies in four strategy sequences with BC-*fi*. Strategies sequenced with BC-*fi*1included elimination of options (*fi*6), rereading of paragraph (*fi*3), use of background knowledge (*fi*8), and matching of content (*fi*7). Among test takers who habitually benefit from sequences that start with BC-*fi*1 was Participant (8), who mentioned that with BC-*fi*: *I read the paragraph and try to understand it first <fi*1> *before the options, because I know that the question will ask about the paragraph especially if the question says "According to the paragraph …*".

Efficacy of strategy use among test takers was associated with the different ways in which strategies were sequenced in strategy clusters. Such forms of strategy sequencing help us describe aspects of effective strategy use among test takers. First, strategy sequences involved strategies that were combined in such a manner that the attached strategy was a modified form of the source strategy. Examples of this type of sequencing include sequences of read and reread strategies (e.g., BC-*ss*1→*ss*6), and strategies of reading local context and those of reading global context (e.g., BC-*v*1→*v*4). Another interesting aspect linked to the first aspect of effective strategy use is that attached strategies served to endorse the functions of source strategies. Source strategies like reading the given paragraph and reading the given sentence were intended to achieve different goals that included figuring out the required information (I-*rp*6) and figuring out

the meaning of the target word (BC-*v*1), respectively. With I-*rp*, the strategy of reading the paragraph (*rp*6) accepted a reread strategy (*rp*3), which allowed test takers more opportunity to infer the required information. And with BC-*v*, the strategy of reading the sentence (*v*1) accepted a strategy which expanded the focus from the local context of the target word to its global context (*v*4). Consequently, test takers who could not figure out the meaning of the target word through the use of strategy (*v*1) were able to achieve this goal by means of strategy (*v*4).

Within certain strategy sequences, source strategies facilitated the use of attached and subsequent strategies. The facilitative role of source strategies extended to make subsequent strategies perform their specific functions in strategy sequences. Test takers found it useful to start responding to certain items by reading the options before reading the paragraphs referred to by these items (e.g., BC-*fi*5→*fi*1, BC-*nf*5→*nf*4, I-*bi*1→*bi*2). With such items, reading the options first served one of two goals that contributed to achieving the purpose of reading the given paragraphs. As most test takers who used this sequence suggested, reading the options first either helped them become familiar with the answer possibilities or provided them with textual elements that hinted at the source of critical information in the paragraph. Participant (21) mentioned that it was his habit to read the options before the paragraph: *I read the options first so that I get less troubled when reading the paragraph as a whole, and I can locate the main points needed for the question.* Other examples of the facilitative role of source strategies in strategy clusters included elimination of options before decision on an option (BC-*v*6→*v*10), and reading of the highlighted sentence before elimination of options (BC-*ss*1→*ss*4).

There were other strategies in strategy clusters that were sequenced in a manner that allowed synchronicity of their use, though as strategy sequencing implies, source strategies were applied first. In fact, all occurrences of synchronized strategies showed that source strategies conduced to the function of attached strategies. For example, when responding to I-*it*, there were test takers who read the paragraph referred to by the I-*it* item (*it*2) and at the same time eliminated those slots where the context sounded irrelevant to the theme of the given sentence (*it*5). Test takers' elimination of certain options (*it*5) was largely dependent on their reading of the paragraph (*it*2). Put differently, the source strategy (*it*2) enabled test takers to apply the attached strategy (*it*5) to the contexts of the slots where the given sentence could be inserted in the paragraph. Other examples of synchronized strategies in strategy clusters included trying the given sentence in one or more of the slots in the paragraph and using a logical connection between the given sentence and the sentence(s) before and/or after the given slot to choose an option (I-*it*4→*it*3), and using understanding of the paragraphs referred to in previous items and eliminating certain answer choices (R2L-*st*3→*st*6).

The strategy of option elimination appeared as an attached strategy in numerous strategy sequences. In most of these sequences, test takers eliminated certain options on the basis of the information they understood from the text. The elimination strategy was attached to source strategies which involved reading a given paragraph or sentence. Therefore, the use of elimination strategy benefited from the grasp of information that test takers had as a result of using the source strategies. This is to say that in almost all of the sequences involving option elimination, the use of this strategy was rational or test-management. In one example, test takers read the paragraph referred to by the I-*bi* item

first (*bi*2), and then eliminated the options that they considered either incorrect or irrelevant (*bi*4) on the basis of the critical information in the paragraph.

Other examples of binary-strategy sequences in which the elimination strategy was inspired by the source strategy include reading the paragraph referred to by the BC-*fi* item (*fi*1), and then eliminating those options that were either incorrect or irrelevant (*fi*6); reading the sentence containing the target pronoun (BC-*pr*1), and then, eliminating those options that were incorrect (*pr*3); reading the highlighted sentence (BC-*ss*1), and then, eliminating those options that were either incorrect or irrelevant in the light of the information content of the highlighted sentence (*ss*4). There were sequences in which the elimination strategy was used almost simultaneously with the source strategy. This observation applies to the use of elimination strategy with I-*it* in two instances. Besides the (*it*2→*it*5) sequence which was illustrated earlier, another sequence involved test takers' trying the given sentence in one or more of the slots in the given paragraph (*it*4) and simultaneously making use of elimination of those slots where the context seemed either distant from the context of the given sentence or irrelevant to its theme (*it*5). In both instances, the source strategy was the first one to be started just shortly before the elimination strategy.

The elimination strategy also appeared in three-strategy sequences. In such sequences, the elimination strategy was inspired by the preceding strategies including the source strategies. In certain sequences, the elimination strategy assumed the second position. One case in point was test takers' use of elimination when responding to BC-*nf*. Test takers read the options first (*nf*5), read the paragraph while eliminating certain options because such options were either incorrect or mentioned in the paragraph (*nf*4),

and then, confirmed the selected answer by rereading the relevant portion of the paragraph (*nf*3). Clearly, test takers' use of elimination in (*nf*5→*nf*4→*nf*3) sequence benefited from their being acquainted with the options and directed their attention to the portion of the paragraph to focus on for the purpose of answer confirmation.

Also, the elimination strategy assumed the third position in three-strategy sequences. One example comes from test takers' response to I-*rp*, in which case test takers read the given paragraph (*rp*6). Then, test takers read the relevant portion of the paragraph more than once (*rp*3) in order to figure out the required information. After that, test takers eliminated those options they found incorrect (*rp*5) based on what they understood. In this sequence, certain options were eliminated on the basis of the information that test takers had learned from their reading the relevant portion of the paragraph more than once. Another example of using the elimination strategy in a three-strategy sequence was test takers' reading of the introductory sentence (*ps*5) on R2L-*ps*, which guided their selection of answer choices that represented important or summary points of the passage (*ps*6). Then, test takers eliminated those answer choices that they considered inappropriate to be summary points (*ps*3). Participant's (4) view that *elimination helps limit the options for consideration* seemed to hold in all instances of using the elimination strategy among test takers.

Among attached strategies in strategy sequences, which were shown to promote effective strategy use were confirmation strategies. As noted earlier, confirmation strategies were mainly used to ensure that the selected answer was the only or most potential option or choice among all available alternatives. In strategy sequences involving confirmation of the selected answer, confirmation strategies often tailed

strategy sequences. Thus, in such instances of strategy sequencing, confirmation strategies functioned to endorse the choice of antecedent strategies and the way they were sequenced. Another facet to strategy sequencing that involved answer confirmation has to do with the manner in which strategies were chosen and used. In three-strategy sequences, the confirmation strategy was either derived from the source strategy or the attached strategy, depending on which strategy was critical to the item response. To illustrate, test takers who confirmed their answers to I-*rp* did so by rereading the paragraph referred to by the I-*rp* item or the relevant portion of it that comprised the sentence containing the highlighted phrase and its surrounding context (*rp*4). In (*rp*1→*rp*2→*rp*4) sequence, strategy (*rp*4) drew on the antecedent strategy (*rp*2), by means of which test takers read the sentence containing the highlighted phrase and the surrounding context to figure out the required information. This is because the attached strategy (*rp*2) was more important to answering I-*rp* than was the source strategy (*rp*1)—the latter being reading the options first.

There were other instances of strategy sequences that ended with confirmation strategies. In the first instance, test takers read the paragraph referred to by the I-*it* item (*it*2), used a logical connection between the given sentence and the sentence(s) before and/or after a given slot to choose an option (*it*3), and confirmed the answer by checking other options or slots to see if any of them accepted the given sentence (*it*6). BC-*nf* provides another example, in which test takers read the options first (*nf*5), read the given paragraph while eliminating certain options (*nf*4), and confirmed the selected answer by rereading the relevant portion of the paragraph (*nf*3).

Another attached strategy shown to promote effective strategy use was deciding on an option. Similar to the strategies of answer confirmation, the strategies of deciding on an option appeared last in their respective strategy sequences. A plausible reason for why strategies of deciding on an option tailed their sequences is that such strategies aided test takers' judgment about which option to pick from two or more options when none of them was certain. In other words, test takers used the strategies of deciding on an option as the last choice to resolve a state of hesitation between two or more options.

In contrast to confirmation strategies which mainly drew on test-management means (e.g., rereading the given paragraph or the relevant portion of it, trying the selected option, checking other options, etc.), strategies of deciding on an option utilized means some of which can be described as test-wise (e.g., semantic clues, guessing, key words/phrases, etc.). For instance, in their response to BC-*v*, test takers eliminated those options they found incorrect or irrelevant (*v*6) and then decided on an option among the ones left by means of semantic clues (*v*10). This illustrates an instance of sequences in which test takers benefited from applying the strategy of elimination before deciding on an option, especially when test takers were left with two or more options after applying the elimination strategy. It is obviously the case that whenever test takers picked an answer, they employed test-management means to confirm the selected answer. On the contrary, whenever test takers were uncertain about which option to choose, they used their test-wiseness to decide on an option.

Additional Analysis

It was hypothesized that high-scoring test takers took full advantage of effective strategy sequencing in their responses to the item types. The assumption here holds that if

high scorers had not made effective strategy use, they would not have obtained high scores. It was postulated that by verifying this hypothesis, effective aspects of strategy use as manifested in the answer to the third research question would gain supporting evidence. Towards this goal, the proportion of effective strategy sequences used by high scorers to the total strategy sequences across all item types was calculated, and found to be almost 75%. By the same token, the proportion of effective strategy sequences used by low scorers was found to equal 17%.

When comparing the results of the two scoring groups, it becomes evident that high scorers surpassed their low counterparts in the use of effective strategy sequences, and consequently in the employment of effective aspects of strategy use. This finding explains high scorers' superior levels of performance and scoring. It also indicates that top scorers enjoyed a high level of awareness of how to use strategies effectively in terms of what strategies to use, and when and how to sequence them. This confirms findings of N. Anderson (1991), Nikolov (2006), Phakiti (2003), and Tian (2000) who noted that the degree of strategic awareness was a major difference between levels of performance and scoring on the language tests used in their studies.

Summary

Chapter IV presented the findings of this research with respect to test-taking strategy use in relation to task and item types, test-taking strategy choice and use in high- and low-scoring groups, and aspects of effective test-taking strategy use across task and item types.

As regards test-taking strategy use in relation to task and item types, the test-taking strategies that test takers used in response to various item types generally

conformed to the demands of item formats. Strategies used by test takers were classified into two groups: Common strategies which were used across a variety of item types, and specific strategies which exhibited a high level of item dependency. Both levels of familiarity with item formats and language proficiency determined the extent to which test takers made use of test-management versus test-wiseness strategies. While test takers used strategies most of which were test-management strategies, they resorted to test-wiseness strategies with challenging items and followed systematic ways in using a variety of technical and textual elements. Strategy use among test takers was purposeful in that test takers applied certain strategies for different goals, multi-form because certain strategies were used by test takers with some variation across several item types, and resourceful since strategies were often applied using different means that were largely determined by item formats and demands in addition to test takers' reading ability and test-taking skill.

As for test-taking strategy choice and use in high- and low-scoring groups, strategies of high scorers were distinct across most of the item types, which points out high scorers' awareness that different item types called for distinct strategies. High scorers made use of strategies that combined aspects of both a high level of understanding of textual information and a superior skill of test-management. They were capable of developing global understanding of the text up to a level that enabled them to answer such item types as R2L-*ps* and -*st* in the same manner intended by the test makers. Top scorers were also disposed to use strategies involving confirmation of the selected answers and checking potential answers, which indicates high levels of strategic awareness and monitoring. They consulted background knowledge in such a manner that

helped them relate their understanding of item content to real-world facts and experiences. On the other hand, low scorers adhered to certain strategies regardless of item formats and demands, and most of these strategies involved the use of clues and guessing. They made use of certain strategies to compensate for deficient comprehension skills, such as rereading the text or part of it or reading the surrounding context, matching of textual elements, replacing vocabulary items with options, and attempting to locate each option in the paragraph with BC-*nf*. Although both high- and low-scoring groups used such strategies as elimination of options, reading of the options before the text, and use of clues, they differed in their intentions behind the use of these strategies.

With regard to aspects of effective test-taking strategy use across item types, it was found that the manner in which test takers sequenced strategies determined the extent to which their strategy use was effective. Source strategies in strategy sequences possessed both a high level of compatibility with item formats and flexibility of accepting other strategies as attached strategies. As determined through strategy sequences, instances of effective strategy use among test takers can be summarized as follows: Certain attached strategies were modified forms of source strategies, certain attached strategies endorsed the functions of source strategies, source strategies facilitated the functions of attached and subsequent strategies, and attached strategies were synchronized with source strategies. There were other aspects of effective strategy use that were prompted by behaviors of specific strategies in strategy sequences. First, the strategy of option elimination was either synchronized with or used after source strategies in order to reduce the options available for consideration. In three-strategy sequences, the strategy of option elimination assumed either the second position to support the function

of a subsequent strategy with BC-*nf*, or the third position to draw on an antecedent strategy with I-*rp*. Second, the strategy of answer confirmation tailed strategy sequences in which it was derived from either a source strategy or an attached strategy, depending on which strategy was critical to the item response. And third, the strategy of deciding on an option tailed strategy sequences in which it served to resolve a state of hesitation between two or more options, as determined by application of an antecedent strategy. It was also found, based on an additional analysis, that the ability to use effective strategy sequences made an important difference between high- and -low scoring groups in favor of the former.

Chapter V provides a summary of the study, conclusions, implications for practice, suggestions for further research, and concluding remarks.

CHAPTER V

CONCLUSIONS

In Chapter IV, the results of this study have been reported, and the findings discussed. This chapter presents conclusions drawn from findings, implications of findings for practice, and suggestions for further research. The chapter is organized into five sections: (a) a summary of the study, (b) conclusions based on findings, (c) implications for practice, (d) suggestions for further research, and (e) concluding remarks.

A Summary of the Study

This section briefly restates the theoretical framework, research questions, methodology, and findings of the study.

Theoretical framework.

***Statement of the problem*.** Previous research has had little to say about how strategy use relates to testing tasks specific to the reading section of the TOEFL iBT. Scholars have called for more research on the relationship among strategy use, test format, and test performance (e.g., Cohen, 2006), with participants who have common backgrounds (e.g., Phakiti, 2008; Rupp et al., 2006). There is also an expressed need for

research into aspects of effective strategy use on the task items on the TOEFL-iBT reading section (Cohen & Upton, 2006). In reply to these calls for further research, this study had as its main goal to explore strategy use on the TOEFL-iBT reading section in a specific population comprising Arab ESL learners.

Purpose of the study. With the focus on how a sample of 25 Arab ESL learners behave in response to the reading tasks on the TOEFL iBT, this research intended to: first, find out what test-taking strategies respondents use when responding to the reading tasks and items; second, investigate if there are differences between high- and low-scorers among respondents in their choice and use of test-taking strategies on the reading tasks and items; and third, determine what aspects of effective test-taking strategy use respondents employ on the reading tasks and items.

Significance of the study. Exploring strategies specific to the TOEFL-iBT reading subtest informs us about how strategy use interacts with test format and performance (cf., Cohen, 2006). Through a study of strategy use among Arab ESL learners when they take the reading subtest of the TOEFL iBT, it is possible to know what strategies are typically used by respondents with the specific task and item types. Secondly, determining differences between high- and low-scoring respondents in strategy use enables us to learn what response behaviors and strategies characterize high test performance, and which act to the contrary. And thirdly, relating strategy use by respondents to their right answers to item types helps us ascertain aspects of effective strategy use on such item types.

The study of respondents' choice and use of test-taking strategies in response to the TOEFL-iBT reading section results in insights about how the population of Arab ESL

learners approaches and performs the reading tasks on the test. Such insights would enable us to propose practical implications for classroom practice, prospective test takers, and test preparation programs. A number of Arab ESL learners whom the researcher has met over the years argued that the reading section of the TOEFL iBT is the most challenging of all four sections on the test. It is hoped that the practical implications of this study can help ease the tension and reduce the difficulty that this population of examinees experiences when taking the reading subtest of the TOEFL iBT.

Research questions.

With the intention to explore strategy use of a sample of 25 Arab ESL learners when they respond to the reading tasks on the TOEFL iBT, this study sought to answer the following questions:

1. What test-taking strategies do subjects use when responding to reading tasks and items?
2. Are there any differences between high- and low-scorers among subjects in their use of test-taking strategies on reading tasks and items?
3. What aspects of effective test-taking strategy use do subjects tend to employ with reading tasks and items?

Methodology.

Data of this study were collected through a procedural integration of stimulated recall, self-observation, and retrospective interview (or SRSORI procedure). Each participant was asked to respond to two reading sets that are part of an actual TOEFL-iBT test administered previously by the ETS. Data collection was carried out over three stages. First, each participant was oriented to the procedures and steps of data collection

and trained in how to produce verbal reports. Second, the participant was engaged in the SRSORI procedure in two sessions separated by a break. And third, the participant was debriefed about his or her performance and scores, and rewarded for participating in the research. Before data collection, a pilot study was conducted in order to check the adequacy of research materials and procedures. On the basis of the outcomes of the pilot study, certain improvements and adjustments were applied to research materials and procedures. For example, the instructions and directions that were part of the orientation and training of participants were made clearer and more elaborate. Also, participants were provided with both a tutorial on how to respond to the reading set and a demonstration of how to use the cursor to express thoughts linked with its movements.

After all collected data were transcribed, verbal reports associated with certain items were selected for coding based on the results of item analyses. A coding scheme was constructed for each item type by means of inductive coding of transcribed data. In data analysis, combining quantitative figures and qualitative accounts of data was deemed ideal to address the three research questions of this study. On the one hand, quantitative figures in terms of frequencies would indicate patterns of strategy use; on the other, qualitative accounts would describe in more detail how strategies relate to such factors as item type, scoring level, and the choice of the right answer.

Findings.

The findings of this research are summarized here with reference to strategy use by task and item type, high- versus low-scorers' strategy use across task items, and aspects of effective strategy use across item types.

Strategy use by task and item type. Strategies used by test takers were classified into two groups: Common strategies which were used across a variety of item types, and specific strategies which exhibited a high level of item dependency. Both levels of familiarity with item formats and language proficiency determined the extent to which test takers made use of test-management versus test-wiseness strategies. While test takers used strategies most of which were test-management strategies, they resorted to test-wiseness strategies with items they found challenging. Test-wiseness strategies involved systematic use of a variety of technical and textual elements. Strategy use among test takers had three facets. First, it was purposeful in that test takers applied certain strategies for different goals. Second, it was multi-form because certain strategies were used by test takers with some variation across several item types. And third, it was resourceful since strategies were often applied using different means that were largely determined by item formats in addition to test takers' reading ability and test-taking skill.

High- versus low-scorers' strategy use across task items. Strategies of high scorers among test takers were dependent upon item format across most of the item types, which points out high scorers' awareness that individual item types call for distinct strategies. High scorers made use of strategies that combined aspects of both a high level of understanding of textual information and a superior skill of test-management. Top scorers were also disposed to use strategies involving confirmation of the selected answers and check of potential answers, which indicates a high level of strategic monitoring. They consulted background knowledge in such a manner that helped them relate their understanding of item content to real-world facts and experiences. On the other hand, low scorers adhered to certain strategies regardless of item formats and

demands, and most of these strategies involved the use of clues and guessing. They made use of certain strategies to compensate for deficient comprehension skills, such as rereading the text or part of it. Although both high- and low-scoring groups used such strategies as elimination of options, reading of the options before the text, and use of clues, they differed in their intentions behind the use of these strategies.

***Aspects of effective strategy use across item types*.** The manner in which test takers sequenced strategies determined the extent to which their strategy use was effective. Source strategies in strategy sequences possessed both a high level of compatibility with item formats and flexibility of accepting other strategies as attached strategies. Aspects of effective strategy use among test takers can be summarized as follows: a) Certain attached strategies were modified forms of source strategies, b) certain attached strategies endorsed the functions of source strategies, c) source strategies facilitated the functions of attached and subsequent strategies, and d) attached strategies were synchronized with source strategies. There were other aspects of effective strategy use that were prompted by behaviors of specific strategies in strategy sequences. First, the strategy of option elimination was either synchronized with or used after source strategies in order to reduce the options available for consideration. In three-strategy sequences, the strategy of option elimination assumed either the second position to support the function of a subsequent strategy with BC-*nf*, or the third position to draw on an antecedent strategy with I-*rp*. Second, the strategy of answer confirmation tailed strategy sequences in which it was derived from either a source strategy or an attached strategy, depending on which strategy was critical to the item response. And third, the strategy of deciding on an option tailed strategy sequences in which it served to resolve a

state of hesitation between two or more options, as determined by application of an antecedent strategy. It was also found, based on an additional analysis, that the ability to use effective strategy sequences made an important difference between high- and -low scoring groups in favor of the former.

Taken together, findings of the current study inform us about how strategy use among 25 Arab ESL learners interacted with item format and performance on the TOEFL-iBT reading tasks. Thus, they address the research goals as expressed in Chapter I. We can now say with certainty that, first, the strategies used by the respondents are typical of them, or are the kinds of strategies they tend to use, when they are engaged in response to various item types on the reading tasks. Second, response behaviors and strategies of a given scoring level are the ones that typify test performance at this level, and vice versa. And third, aspects of effective strategy use among the respondents are those that increase their chance of getting items right on the reading tasks. As such, findings of this study break new ground for exploration of strategy use on the TOEFL-iBT reading section, among respondents who considerably share linguistic and cultural backgrounds.

Findings contribute to the knowledge and research bases of language testing and assessment, specifically proficiency-oriented testing. This contribution pertains to three converging areas of strategy use among Arab ESL learners on the reading section of the TOEFL iBT. First, test takers use strategies depending on item format and difficulty in ways that allow them to achieve different goals, adapt certain strategies to various item types, and apply strategies using several textual and technical means. Second, high test performance and scoring are characterized by superior skills of both comprehension and

test-management as well as high levels of strategic awareness and monitoring. Conversely, low test performance and scoring are associated with poor skills of comprehension and excessive use of test-wiseness. Third, test takers sequence strategies such that certain strategies derive from other strategies, endorse or facilitate the functions of other strategies, or act in sync with other strategies. Such strategy sequences were linked with other aspects of effective strategy use that included reducing the options available for consideration, confirming the choice of the right answer, and deciding on an option among two or more options.

Conclusions

The conclusions from this study are addressed here such that they align with the research questions and findings.

Strategy use by task and item type.

The kinds of test-taking strategies that examinees use on the TOEFL-iBT reading tasks are determined by item formats. The fact that item types on the reading tasks differ in their formats demands that test takers use strategies that are tailored to item formats taken individually. N Anderson et al. (1991) arrived at the same conclusion in a study of strategy use on the reading section of the TOEFL-PBT. It goes without saying that the more familiar test takers are with item formats the more likely that they will use strategies that help them manage their response to item types, or the so-called test-management strategies. Therefore, we can say that test-management strategies used with item types are necessarily compatible with the formats of these items.

Besides task familiarity, another factor that influences test takers' decisions as to what strategies to use is the level of language proficiency. Language proficiency relates

closely to the distinction between test-management and -wisness strategies. It seems that high proficiency prompts use of test-management strategies through what Mangubhai (1991) referred to as resourcefulness which denotes efficient and automatic use of mental and textual resources on testing tasks. Test takers resort to test-wiseness strategies when their levels of proficiency do not enable them to choose the right answer. On tests of reading comprehension, the use of test-wiseness strategies can also be due to text complexity and task difficulty (Cohen & Upton, 2006). It follows that test takers whose proficiency levels are high intermediate and above, or similar to the levels of the high scorers' in this study, use more test-management strategies than test-wiseness strategies. The use of test-wiseness strategies among high-proficiency test takers is limited to their encounter with challenging test items.

Both reading proficiency and test-taking skill determine the extent to which test takers can benefit from three facets of strategy use on the TOEFL-iBT reading tasks. In other words, test takers with high levels of reading proficiency and test-taking skill can use certain strategies with various item types to, first, achieve more than one goal. For example, reading the sentence containing the target word with BC-*v* can be aimed to either figure out the word meaning from the context or confirm the known meaning of the word. Second, test takers may use variant forms of certain strategies across several item types. For example, test takers can choose to read the paragraph referred to by a given item or a portion of it, or the context specified by the item to extract the answer. And third, test takers can use different means in their application of certain strategies. For example, the strategy of confirming the selected answer can be applied by rereading the given paragraph or a portion of it, or checking options other than the chosen one.

Strategies used by test takers on the TOEFL-iBT reading tasks can be fitted into two groups. The first group comprises common strategies which can be used across a variety of item types. Thus, such strategies are highly compatible with, or adaptable to, several item formats. Also, these strategies are not dependent on the nature of the reading task; therefore, they are functional with the three task types on the reading subtest, namely basic comprehension, inferencing, and reading-to-learn. Examples of such strategies include eliminating certain options, confirming the selected answer, and deciding on an option among two or more options. The second group of strategies consists of strategies that are unique or specific to certain item types. These strategies are used more frequently compared to common strategies, which is due to the fact that most item types on the reading tasks have their distinct formats and demands. Examples of such strategies include attempting to locate each option in the given paragraph while eliminating certain options with BC-*nf*, trying the given sentence in one or more of the slots in the given paragraph with I-*it*, and selecting answer choices that represent important or summary points of the passage with R2L-*ps*. Rupp et al. (2006) used similar labels to describe two groups of strategies that their respondents used on reading tests for academic purposes.

High- versus low-scorers' strategy use across task items.

Responses and performances of the two discrete levels of scoring on the TOEFL-iBT reading tasks are linked with the use of strategies that set these levels apart. High scorers are exceptionally aware that different item types on the reading tasks call for distinct strategies that match formats and demands of these items. As mentioned earlier, this trait characterizes strategy use of high-proficiency test takers, because scoring level

on the reading tasks correlates with reading proficiency. Such awareness of what strategies to use with which question items increases with proficiency level, and so augments test performance (N. Anderson, 1991; Nikolov, 2006; Tian, 2000). High performance and scoring on the reading tasks also require adequate grasp of textual information. The abilities to develop global understanding and create a mental schema of the main points of the text are characteristics of high scorers. Such abilities assist test takers in their response to R2L-*ps* and -*st* in a manner consistent with the test constructors' intentions (ETS, 2009). Hence, high scores on reading tasks necessitate that examinees be proficient comprehenders and skilled test takers.

In addition to strategic awareness, high performance and scoring on the TOEFL-iBT reading tasks profit from strategies that involve confirming the selected answers and checking potential answers. The extent to which test takers can make use of these strategies relies on their abilities of reading comprehension. Also, the use of these strategies seems to correlate with the abilities to assess and monitor strategy use. Such abilities were shown to favor high scorers in previous studies (e.g., N. Anderson, 1991; Nikolov, 2006; Tian, 2000). High performance and scoring on reading tasks draw on background knowledge such that high scorers can relate their understanding of textual information to real-world facts and experiences. More specifically, high scorers are able to combine their understanding of the text with real-life knowledge and/or experience in response to such task items as BC-*v*, BC-*fi*, BC-*ss*, and I-*rp*. Consulting background knowledge and experience is considered a metacognitive strategy which allows test takers to confirm their answers (Nikolov, 2006).

In contrast to the aspects that feature high performance and scoring, low performance and scoring are typical of test takers who have insufficient strategic awareness of strategy-format compatibility. However, it is possible that these test takers are aware of what strategies to use with various item types, but they exhibit some deficiency when it comes to the manner in which strategies ought to be used. In other words, test takers who are not aware of what strategies to use with what item formats or how to use these strategies are not likely to perform well on reading tasks and attain high scores. Indeed, low performance and scoring are associated with the use of certain strategies regardless of item formats and demands, or what Tian (2000) described as rigidity of strategy use.

Besides, strategies that involve use of clues and guessing make up most of the strategies that low scorers tend to employ with reading tasks. Such strategies are test-wise because they utilize textual and/or technical aspects of the test that are not construct-relevant, or follow ways that run counter to test constructors' intentions (Allan, 1992; Cohen, 2006; Cohen & Upton, 2006, Phakiti, 2008). Therefore, the assumption made earlier that test takers who are at a low level of proficiency are disposed to use more test-wiseness strategies than test-management strategies is supported. Low scorers use strategies to compensate for deficient comprehension skills. For the most part, such strategies call for reading at the local level of the text and may embrace test-wiseness, for example, matching the content of the text or elements of it and the options. However, when the need arises, low scorers may try to extend their reading focus beyond the context indicated by the item, especially with BC-*v* and -*pr*. Reading the surrounding

context with BC-*v* and -*pr* allows test takers to attend to more of the text, which can help them grapple with the choice of the answer.

Last, high and low scorers may appear to be using the same strategies; however, their intentions of using such strategies may be totally different. As a general rule, this reflects aspects of test-management versus -wiseness strategy use. To be specific, high scorers use strategies in a manner that helps them arrive at the answer by following a meaningful and systematic route. Low scorers, on the other hand, may use the same strategies with their aim being to compensate for their deficient ability, or lack thereof, to get to the answer. For example, high scorers may choose to read the options before the text to have familiarity with the answer possibilities. Conversely, low scorers may use the same strategy in order to use the options to locate the source of critical information in the text. Other examples of strategies that can be used by the two discrete levels of scoring, yet for different reasons include eliminating certain options and using clues. It is interesting to note that these two strategies, as well as reading the options first, belong to the category of common strategies that can be applied to a variety of item types.

Aspects of effective strategy use across item types.

Aspects of effective strategy use on the TOEFL-iBT reading tasks can be deduced from the manner in which test takers cluster strategies in response to various item types. The degree to which strategies are logically sequenced within strategy clusters makes a difference between high and low test performance. Logical sequences of strategies often stem from source strategies that are both highly compatible with item formats and acceptable of attached strategies. Source strategies in strategy sequences should be intrinsically relevant to the nature of the task being performed (N. Anderson, 2005).

Strategies in a sequence also need to be in harmony with one another (Nikolov, 2006). Test takers who make use of effective strategy sequencing are more likely to perform exceptionally well and score high on reading tasks than those who do not.

Forms of logical sequencing of strategies enable us to discern aspects of effective strategy use that reflect the role of individual strategies within strategy sequences. First, attached strategies can be modified forms of source strategies, in which case attached strategies endorse the functions of source strategies. Second, source strategies may facilitate the use of attached and subsequent strategies, in which case the specific functions of strategy sequences cannot be attained without the facilitative role of source strategies. And third, source strategies are synchronized with attached strategies such that both function almost simultaneously to help achieve the goal of the strategy sequence as a whole. Although certain strategies can be synchronized in strategy sequences, source strategies are applied shortly before attached strategies with respect to time frame. The way synchronized strategies are sequenced suits the role of source strategies in that they act as conducive to, or they set the stage for, the function of attached strategies.

Aspects of effective strategy use on reading tasks can be prompted by the specific functions that individual strategies perform in strategy sequences. Such strategies belong to the group of common strategies that are applicable across item types and can be used in two- and three-strategy sequences. These strategies are eliminating certain options, confirming the selected answer, and deciding on an option between two or more options. The strategy of eliminating certain options can either be synchronized with or used following the source strategy. In both cases, the strategy performs its function by helping test takers reduce the options available for consideration. The use of elimination strategy

is rational or test-management so long as it benefits from a grasp of textual information, or is inspired by an antecedent strategy which calls for reading of the text.

The strategy of confirming the selected answer tails strategy sequences. The confirmation strategy is intended to ensure that the selected answer is the only or the most potential option or choice among all available alternatives. In two-strategy sequences, the confirmation strategy is often a modified form of the source strategy. In three-strategy sequences, the confirmation strategy can be derived from the source strategy or an antecedent strategy, whichever is shown to be critical to the choice of the answer. The strategy of deciding on an option appears last in strategy sequences to help resolve a state of hesitation between two or more options. In contrast to confirmation strategies which mostly draw on test-management means, strategies of deciding on an option may use means some of which are test-wise. It is plausible to say that whenever test takers want to confirm their answer, they are likely to use test-management means (e.g., rereading the text). But, when test takers are not certain about which option or choice to select, they are likely to employ test-wiseness means (e.g., clues).

Generally, as far as strategy use on the TOEFL-iBT reading tasks is concerned, strategies are of two types: common strategies that can be used with several item types, and specific strategies whose use is linked with specific item types. High performance and scoring on reading tasks draw on strategies that differ from those associated with low performance and scoring. Effective aspects of strategy use occur in tandem with logical sequences of strategies.

Implications for Practice

The implications of this study for practice are grounded in the findings and conclusions. In general, implications are tailored to the needs of prospective test takers of the TOEFL iBT. As mentioned in Chapter I, the implications here are related to classroom practice, test takers, and test preparation programs. Although the implications here are for ESL learners, they are also assumed to be practicable in contexts where English is taught or learned as a foreign language.

Classroom practice.

We are living in an era that is witnessing a change from the tradition that upheld the belief that testing and assessment are made for their own sake. This change is in the direction of making testing and assessment on the one hand, and teaching and learning on the other hand feed into each other. The expectations are high that the TOEFL iBT in general will have positive washback, or will influence teaching and learning practices in constructive ways. For example, Bejar, Douglas, Jamieson, Nissan, and Turner (2000) argued that the content offered in contexts of English for academic purposes will become increasingly communicative. Based on their studies of how the TOEFL iBT impacts English language teaching and learning in a number of countries in Europe, Wall and Horák (2006, 2008a, 2008b) concluded that teachers have started to incorporate more speaking and writing tasks into their lessons, following an integrated approach. Also, teachers are now focusing on skills of analytical and integrated reading more than ever before.

Any research effort in the field of language testing and assessment should have implications for classroom practice so that whatever is learned from research contributes

to the way we offer language classes. The implications for classroom practice aim to get ESL learners to develop skills of reading comprehension that characterize high performance and scoring on the TOEFL-iBT reading tasks. Instructors are advised to guide their ESL students to develop reading skills in consideration of the interactive view of reading comprehension. According to this view, language teaching and learning should pursue a combination of both categories of bottom-up and top-down processes in the development of reading skills (Hudson, 1996). Learners need to be able to perform the two modes of bottom-up and top-down processing in order to respond properly to R2L task items on the TOEFL iBT.

Learners should be instructed and trained in how to develop global understanding of the text and how to create a mental schema of its major points. An exemplary lesson with this aim has three components: First, learners are provided with instructions and guidance about what these skills involve. Second, instructors demonstrate to learners how to apply these skills using an overhead or PowerPoint display. And third, learners are offered ample practice with reading texts to enable them to automatize and master these skills. Readers ought to be encouraged to intertwine their background knowledge and experience with their understanding of textual information. This technique helps readers endorse and/or evaluate their understanding of the text.

When learners are engaged in reading test-taking, they ought to be able to combine understanding of textual information and use of strategies that can help them regulate their test performance. Certainly, this would require instructors to provide learners with reading tasks that resemble those on the TOEFL-iBT reading section in terms of task and item formats. Instructors should also offer necessary training in the use

of test-taking strategies that are compatible with task and item formats. It is advisable that learners be trained in how to check potential answers and confirm the selected answers on the reading test-taking tasks. There are two potential outcomes of this strategy training: First, learners can achieve a level of strategic awareness and monitoring that would enable them to perform well and achieve high scores on the TOEFL-iBT reading tasks. Second, learners can equip themselves with skills of academic reading. More implications for test preparation that pertain to classroom practice are discussed in the sections below.

Test takers.

Findings and conclusions of this study offer numerous insights for prospective test takers of the TOEFL iBT that pertain to the reading section of the test. The following implications are intended as recommendations for test takers who prepare on their own, or get ready for the test through self-preparation. The first insight comes from the fact that item types on the TOEFL-iBT reading tasks differ in their formats, which demands that test takers familiarize themselves with the formats of the various task items. It is by means of intensive practice with item formats that test takers will be able to develop strategies that would enable them to manage their response to item types, or test-management strategies. In order for test takers to develop these strategies, they should first follow the instructions of the reading section and the specific directions of item types.

It is recommended that test takers acquaint themselves with the instructions that appear on the first screen of the reading section, so that they can only skim through them on the test day (cf., Forster et al., 1997). Also, test takers should acquaint themselves with

the directions specific to each item type, so that they do not have to read them when they take the official test. It would save them much time if test takers can dispense with reading both the instructions and the directions on the test day, because once they start a reading set the allotted time runs. This is especially important because the last items on reading sets, or R2L task items, typically demand much more time than any other item type. Towards the goal of saving more time on the test day, test takers are advised to read the instructions and directions and apply them a number of times in their preparation for the test. It is also advisable that test takers practice using the allotted time in proportion to the number of items, so that no item takes more time than others.

An important consideration in test preparation is that test takers should not feel limited to suggestions or tips about strategies to use in response to the TOEFL-iBT reading tasks. Publishers of test preparation materials recommend that test takers use certain strategies and not use others. Instead, test takers are advised to consider these recommendations as providing them with options from which they can choose certain strategies and experiment with these strategies to decide whether or not to adopt them. Test takers should also allow themselves time to develop and adapt strategies to suit their cognitive and linguistic abilities. Once a test taker knows what strategies to use on various task items, he can be said to have achieved the level of strategic awareness that may boost his test performance and score.

Prospective examinees should be aware that test-taking strategy use makes a difference in performance and scoring on the reading section of the TOEFL iBT. It should also be clear that the ETS encourages the use of test-taking strategies even at the level of individual items on the test, lest examinees think that by using strategies they are

breaking ethical codes or violating principles of academic integrity. In their preparation for the reading section of the test, examinees should try to combine their understanding of textual information and use of test-taking strategies in a series of practice. This combination needs to be exercised over an extended period of time in order for it to be automatic. Intensive and sustained practice helps moderate test anxiety and speed up the response process (Tian, 2000). The preparation for the test should be made using TOEFL-iBT practice tests that are available via the TOEFL website or marketed by publishers of test preparation materials.

Test takers ought to know that they can apply two sets of strategies on the TOEFL-iBT reading tasks. The first set comprises strategies that can be used across almost all item types, or common strategies. This group of strategies includes eliminating certain options, confirming the selected answer, and deciding on an option between two or more options. And the second set consists of strategies that are specific to individual item types. This set of strategies represents strategies whose use is linked to item format, for example, reading the sentence containing the target word with BC-*v*, trying the given sentence in one or more of the slots in the given paragraph to see where it best fits with I-*it*, and selecting answer choices that represent important or summary points of the passage with R2L-*ps.* This knowledge should become part of the strategic awareness that test-takers should develop in their preparation for the test.

Test takers should know that, first, strategies on the TOEFL-iBT reading tasks can be used to achieve more than one goal. For example, the strategy of rereading the paragraph or a portion of it can be used to improve understanding of textual information, locate the source of critical information, or check the adequacy of an option. Second,

variant forms of strategies can be used across item types; for example, test takers can read the whole paragraph, read a portion of it, or skim the paragraph. And third, it is possible to use different means to apply various strategies; for example, test takers can confirm their selected answer by rereading the given paragraph or a portion of it, checking the other possible answers, or ensuring elimination of the incorrect options. It is assumed that getting oneself acquainted with these three facets of strategy use on reading tasks would enable the test taker to capitalize on her cognitive and linguistic abilities.

When test takers master strategies that are highly compatible with item formats, they can use these strategies as source strategies in sequences with other strategies. Source strategies represent strategies from which strategy sequences stem, or the first strategies in strategy sequences. Logical sequencing of strategies would enable test takers to benefit from aspects of effective strategy use on the TOEFL-iBT reading tasks. Strategies in logical sequences can endorse or facilitate functions of one another. They can also be synchronized with one another. Such roles within a given strategy sequence are essential for attaining the goal of the strategy sequence as a whole. Intensive and sustained practice has an important role in getting test takers to use logical sequences of strategies. Also, the manner in which certain strategies are sequenced determines the extent to which these strategies promote effective strategy use through their specific functions. For example, the strategy of eliminating certain options can be more effective when inspired by a preceding strategy which calls for reading. The strategy of confirming the selected answer should appear last in strategy sequences so that it can perform its function properly. The strategy of deciding on an option between two or more options performs its function best if used at the end of the strategy sequence.

Test preparation programs.

Programs that aim to prepare ESL learners to take the TOEFL iBT are advised to base their test preparation practices on research that ETS and specialists in language testing have conducted. Our practices in test preparation programs should not be pursued on the basis of our hunches about strategies that may help test takers perform better and score higher on a given test. Approaches to test preparation not substantiated by research in the field of language testing and assessment suffer two obvious weaknesses: First, they lack empirical evidence which explains why test takers should use certain strategies and not others. And second, they seem to imply that test takers are all equal in cognitive processing (e.g., working memory) and language abilities. Thus, recommendations about strategy use on language tests should originate from empirical investigations and should be flexible enough to accommodate individual differences in cognitive processing and language abilities. One way to cater for individual differences is to inform test takers about "the broad range of strategy options available to them" (Yien, 2001, p. 39), which would allow them to try among these options and be able to decide on whatever option suits their cognitive and linguistic abilities.

As N. Anderson et al. (1991) recommended for the case of the TOEFL-PBT, instructors in programs of test preparation for the TOEFL iBT should set as their main goal training of prospective test takers in how to respond to item types to the best of their language proficiency. This suggests that test preparation be implemented in such a manner that minimizes the possibility of any intervening factors that may bias test takers' scores on the actual test. Such a factor as lack of familiarity with item formats may exercise a construct-irrelevant effect on test performance and score. It should be

emphasized at the beginning of the test preparation course that effective use of strategies depends on how strategies are combined with other strategies and how strategy combinations serve test performance. As the course proceeds, test preparation should incorporate instruction, demonstration, and practice in order that test takers master the skills that are linked with effective test-taking strategy use.

Programs of test preparation for the TOEFL iBT may follow most of the implications for test takers who practice self-preparation as described above. These include, but are not limited to, the following. Test takers should be familiarized with formats of various item types on reading tasks. This would require the provision of TOEFL-iBT practice tests. Such materials are available for download via the official TOEFL website (at http://www.ets.org/toefl/) or through booksellers of test preparation materials. Test takers should be encouraged to get used to the instructions of the reading section and the specific directions of item types. It is highly recommended that test takers be presented with various choices of what strategies to use and how to use these strategies, and be allowed to make the choice that would presumably suit their idiosyncratic characteristics. For example, test takers should be allowed to make their choice of whether to start with the text or the options while being informed about the usefulness of each strategy. To be specific, test takers may be told that the former strategy enables one to develop global understanding of the text, whereas the latter helps one become familiar with answer possibilities.

Instructors are advised to inform test takers that two strategy sets can be used on the TOEFL-iBT reading section, which comprise common strategies and specific strategies as discussed above. Also, it is advisable that instructors get test takers to know

and practice the ways in which strategies can be used on reading tasks to achieve more than one goal with a single item type, modified so as to be usable across several item types, or applied using various textual and/or technical means. Once more, combinations of instruction, demonstration, and practice are necessary to have examinees become aware of these features of strategy use and be able to make use of them.

When practicing with reading sets, examinees should be urged to use the allotted time in proportion to the number of items, and leave enough time to work with R2L task items. Examinees should be advised to do their best to combine their understanding of textual information and use of test-taking strategies. Getting examinees to routinize such response behaviors and strategies calls for intensive and sustained practice. Therefore, constant evaluation and provision of assistive feedback are of primary importance to ensure that the course objectives are met. Part of strategy training ought to be devoted to guiding examinees to make effective use of strategies, first, by using strategies that are compatible with item formats, and second, by following these strategies with strategies such that this sequence of strategies would be logical and beneficial.

Instructors are advised to present trainees with various possibilities of using strategies in logical sequences and impart to them when a certain sequence is favored over another. This form of training should include the effective uses and sequences of the three common strategies, namely eliminating certain options, confirming the selected answer, and deciding on an option between two or more options. Last but not least, care ought to be taken to combine instruction, demonstration, practice, evaluation, and assistive feedback to ensure that examinees can make the most of the test preparation course.

Suggestions for Further Research

Future research into test-taking strategy use on the TOEFL-iBT reading section is recommended to employ the research design of the current study with different language groups. This would serve to extend this research to other contexts and populations of ESL learners. It is likely that some, if not all, of the findings and conclusions reached in this study would prove to be generalizable to other populations of ESL learners. Through further studies with different populations, it would be possible to determine whether or not L1 has any effects on test-taking strategy use on the TOEFL-iBT reading tasks. Accordingly, this may open the door to another area of enquiry into how typological differences among different L1s affect test-taking strategy use on the TOEFL-iBT reading section.

Future studies may attempt to improve this research by recruiting a larger sample of Arab ESL learners and using random sampling in participant selection. This would permit the use of inferential statistics to determine generalizability of findings and conclusions to the entire population of Arab ESL learners. Future researchers are encouraged to make use of the procedural integration followed in this study to explore strategy use on the other sections of the TOEFL-iBT including listening and writing subtests. The integration of stimulated recall, self-observation, and retrospective interview (or SRSORI technique) in this study furnished rich and high quality data, which suffices to encourage the use of this technique to study strategy use on testing tasks in general.

Further research is needed to contribute to our knowledge base of whether strategy use on the TOEFL-iBT reading tasks influences test performance and scores

through causality. In other words, it would be interesting to investigate if there exists a causal relationship between strategy use and test performance as determined by scores. Another avenue of research can be aimed to evaluate the effectiveness of test preparation programs. This can be done through the use of a pretest-posttest design according to which the researcher makes a comparison between test performance before and after test takers receive a course in test preparation. Future researchers may consider using analytical techniques with strategy data to answer numerous research questions. For example, regression analyses can be used to examine the causal relationship between using certain strategy sequences and getting the right answers to item types on reading tasks.

Concluding Remarks

"There are two ways of spreading light: to be the candle or the mirror that reflects it."

-EdithWharton

In writing the last section of the final chapter of my dissertation, I hope to have contributed to the knowledge and research bases of the TOEFL iBT in specific, and the field of language testing and assessment in general. I hope to have lit, or added some light to, the way for prospective test takers and programs of test preparation towards accomplishing their goals pertaining to the reading section of the TOEFL iBT. This has been my goal as a researcher since the moment my research advisor and I agreed about research that I could conduct in connection with Arab ESL learners' dealing with the TOEFL iBT.

It was of particular interest to me to see that my procedural integration of stimulated recall, self-observation, and retrospective interview was so fruitful. I was able

to collect data that were as rich and high-quality as I hoped for. Participants were cooperative and understanding of my situation as a researcher. I greatly appreciate their keenness and commitment throughout data collection. Results and findings were far more interesting than what I imagined at the outset. I learned that test-taking strategy use is more ingenious and versatile than ever described in the literature. I found support for the idea that systematic ordering of strategies can be related to the choice of the right answer.

Overall, this study was such a rewarding and enriching experience that it will keep driving me to pursue further research into how Arab ESL learners deal with the TOEFL iBT. I had the experience of taking the TOEFL-CBT, and I knew how test-taking strategy use is crucial to scoring high on the test. Arab ESL learners struggle to attain scores on the TOEFL iBT that would allow them to study at English-medium universities. They consider the reading section to be the most challenging of all sections on the test. Now, I am assuming the role of a researcher who looks forward to assisting prospective test takers among Arab ESL learners make the most of strategy use on the TOEFL iBT.

REFERENCES

Abanomey, A. (2002). *The effect of texts' authenticity on reading-comprehension test-taking strategies used by adult Saudi learners of English as a foreign language* (Doctoral dissertation). Available from ProQuest Dissertation and Theses database (UMI No. 3069767)

Aebersold, J. & Field, M. (1997). *From reader to reading teacher*: *Issues and strategies for second language classrooms*. Cambridge, UK: Cambridge University Press.

Afflerbach, P. (2002). Verbal reports and protocol analysis. In M. Kamil, P. Mosenthal, P. Pearson, & R. Barr (Eds.), *Methods of literacy research: The methodology chapters from the handbook of reading research, Volume III* (pp. 87-103). Mahwah, NJ: Lawrence Erlbaum Associates.

Afflerbach, P., & Johnston, P. (1984). On the use of verbal reports in reading research. *Journal of Reading Behavior, 16,* 307-322.

Afflerbach, P., Pearson, P., & Paris, S. (2008). Skills and strategies: Their differences, their relationships, and why it matters. In K. Mokhtari & R. Sheorey (Eds.), *Reading strategies of first- and second-language learners: See how they read* (pp. 11-24). Norwood, MA: Christopher-Gordon.

Alderson, J. (2000). *Assessing reading*. Cambridge, UK: Cambridge University Press.

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

Alderson, J. (2009). Test review: Test of English as a Foreign Language: Internet-based Test (TOEFL iBT). *Language Testing, 26*(4), 621-631.

Alderson, J., & Lukmani, Y. (1989). Cognition and reading: Cognitive levels as embodied in test questions. *Reading in a Foreign Language, 5*, 353-370.

Alexander, P., & Jetton, T. (2000). Learning from text: A multidimensional and developmental perspective. In M. Kamil, P. Mosenthal, P. Pearson, & R. Barr (Eds.), *Handbook of reading research: Volume III* (pp. 285-310). Mahwah, NJ: Lawrence Erlbaum Associates.

Alfassi, M. (2004). Reading to learn: Effects of combined strategy instruction on high school students. *The Journal of Educational Research, 97*(4), 171-184.

Allan, A. (1992). Development and validation of a scale to measure test-wiseness in EFL/ESL reading test takers. *Language Testing, 9*(2), 101-122.

Allan, A. (1995). Begging the questionnaire: Instrument effect on readers' responses to a self-report checklist. *Language Testing, 12*, 133-153.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education.(1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anderson, N. (1991). Individual differences in strategy use in second language reading and testing. *Modern Language Journal, 75*(4), 460-472.

Anderson, N. (2005). L2 learning strategies. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 757-772). Mahwah, NJ: Lawrence Erlbaum Associates .

Anderson, N., Bachman, L., Perkins, K., & Cohen, A. (1991). An exploratory study into the construct validity of a reading comprehension test: Triangulation of data sources. *Language Testing, 8*(1), 41-66.

Anderson, R., & Pearson, P. (1988). A schema-theoretic view of basic processes in reading comprehension. In P. Carrell, J. Devine, & D. Eskey (Eds.), *Interactive approaches to second language reading* (pp. 37-55). Cambridge: Cambridge University Press.

Ark, L., Croon, M., & Sijtsma, K. (2005). *New developments in categorical data analysis for the social and behavioral sciences*. Quantitative methodology series. Mahwah, NJ: Lawrence Erlbaum Associates.

Assessment Systems Corporation. (2010). Iteman 4 (Version 4) [Computer software]. St. Paul, MN: Author.

Bachman, L. (1985). Performance on cloze tests with fixed-ratio and rational deletions. *TESOL Quarterly, 19*(3), 535-556.

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

Bachman, L. (1991). What does language testing have to offer? *TESOL Quarterly, 25*, 671-704.

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

Baker, L. (2008). Metacognitive development in reading: Contributors and consequences. In K. Mokhtari & R. Sheorey (Eds.), *Reading strategies of first- and second-*

language learners: See how they read (pp. 25-42). Norwood, MA: Christopher-Gordon.

Baker, L., & Brown, A. (1984). Metacognitive skills and reading. In P. Pearson (Ed.), *Handbook of reading research* (pp. 353-394). New York, NY: Longman.

Balota, D. (1990). The role of meaning in word recognition. In D. Balota, G. d'Arcias, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 9-32). Hillsdale, NJ: Lawrence Erlbaum Associates.

Barnett, M. (1988). Reading through context: How real and perceived strategy use affects L2 comprehension. *Modern Language Journal, 72*, 150-162.

Baumann, J., Jones, L., & Seifert-Kessell, N. (1993). Using think alouds to enhance children's comprehension monitoring abilities. *The Reading Teacher, 47*(3), 184-193.

Beck, I., & McKeown, M. (1989). Expository text for young readers: The issue of coherence. In L. Resnick (Ed.), *Knowing, learning, and instruction* (pp. 47-65). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bejar, I., Douglas, D., Jamieson, J., Nissan, S., & Turner, J. (2000). *TOEFL 2000 listening framework: A working paper*. (TOEFL Monograph Series NO. MA-19). Princeton, NJ: ETS.

Bennett, R. (1993). On the meanings of constructed response. In R. Bennett & W. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment* (pp. 1-27). Hillsdale, NJ: Lawrence Erlbaum Associates.

Biber, D., Conrad, S., Reppen, R., Byrd, P., Helt, M., Clark, V., et al. (2004). *Representing language use in the university: Analysis of the TOEFL 2000 spoken and written academic language corpus* (TOEFL Monograph No. MS-25). Princeton, NJ: ETS.

Birch, B. (2002). *English L2 reading: Getting to the bottom*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bland, J. (1995). *How adult readers navigate through expository text in a hypermedia environment to construct meaning* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9612595)

Block, E. (1986). The comprehension strategies of second language readers. *TESOL Quarterly, 20*, 463-494.

Bloom, B. (1954). The thought processes of students in discussion. In S. French (Ed.), *Accent on teaching: Experiments in general education* (1st ed.) (pp. 23-46). New York, NY: Harper & Brothers.

Blueberry Software.(2009). BB FlashBack Express (Version 2.7.1) [Computer software]. Birmingham, UK: Author.

Bordens, K., & Abbott, B. (2005). *Research and design methods: A process approach*. Boston, MA: McGraw-Hill.

Bower, G., Black, J., & Turner, T. (1979). Scripts in memory for text. *Cognitive Psychology, 11,* 177-220.

Britt, M., & Gabrys, G. (2001).Teaching advanced literacy skills for the World Wide Web. In C. Wolfe (Ed.), *Learning and teaching on the World Wide Web* (pp. 73-90). San Diego, CA: Academic Press.

Brown, A., & Day, J. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior, 22*(1), 1-14.

Brown, A., Iwashita, N., & McNamara, T. (2005). *An examination of rater orientations and test-taker performance on English for academic purposes speaking tasks* (TOEFL Monograph No. MS-29). Princeton, NJ: ETS.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96,* 31-42.

Callies, M. (2009). *Information highlighting in advanced learner English: The syntax-pragmatics interface in second language acquisition*. Amsterdam: John Benjamins Publications.

Carpenter, P., Mikaye, A., & Just, M. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp.1075-1122). San Diego, CA: Academic Press.

Carrell, P. (1984). The effects of rhetorical organization on ESL readers. *TESOL Quarterly, 17*, 441-469.

Carrell, P. (1985). Facilitating ESL reading by teaching text structure. *TESOL Quarterly, 19*, 727-752.

Carrell, P. (1989). Metacognitive awareness and second language reading. *Modern Language Journal, 73*, 121-134.

Carver, R. (1997). Reading for one second, one minute, or one year from the perspective of rauding theory. *Scientific Studies of Reading, 1*(1), 3-43.

Chamot, A. (2005). The role of learning strategies in second language acquisition. In K. Pitt (Ed.), *Debates in ESOL teaching and learning: Culture, communities and classrooms* (pp. 55-64). Harlow, UK: Longman.

Chapelle, C., Enright, M., & Jamieson, J. (Eds.) (2008). *Building a validity argument for the Test of English as a Foreign Language*. New York, NY: Routledge.

Clarke, M. (1979). The short circuit hypothesis of ESL reading or when language competence interferes with reading performance. *Modern Language Journal, 64,* 203-209.

Cohen, A. (1984). On taking language tests: What the students report. *Language Testing, 1*(1), 70-81.

Cohen, A. (1986). Mentalistic measures in reading strategy research: Some recent findings. *The ESP Journal, 5*(2), 131-145.

Cohen, A. (1992). Test-taking strategies on language tests. *Journal of English and Foreign Languages, 10-11*, 90-105.

Cohen, A. (1994a). *Assessing language ability in the classroom* (2nd ed.). Boston, MA: Newbury House/Heinle & Heinle.

Cohen, A. (1994b). English for academic purposes in Brazil: The use of summary tasks. In C. Hill & K. Parry (Eds.), *From testing to assessment: English as an international language* (pp. 174-204). London, UK: Longman.

Cohen, A. (1996). Verbal reports as a source of insights into second language learner strategies. *Applied Language Learning, 7,* 5-24.

Cohen, A. (1998). Strategies and processes in test taking and SLA. In L. Bachman & A. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 90-111). Cambridge, UK: Cambridge University Press.

Cohen, A. (2005). *Coming to terms with language learner strategies: What do strategy experts think about the terminology and where would they direct their research?* Working Paper No. 12. Research Paper Series. Auckland, NZ: Centre for Research in International Education, AIS St. Helens. Retrieved May 23, 2009, from http://www.crie.org.nz/research_paper/Andrew%20Cohen%20WP12.pdf

Cohen, A. (2006). The coming of age of research on test-taking strategies. *Language Assessment Quarterly, 3*(4), 307-331.

Cohen, A., & Aphek, E. (1979). *Easifying second language learning.* Jerusalem, Israel: The Jacob Hiatt Institute. (ERIC Document Reproduction Service No. ED163753). Retrieved September 23, 2009, from ERIC database.

Cohen, A., & Hosenfeld, C. (1981). Some uses of mentalistic data in second language research. *Language Learning, 31*, 285-313.

Cohen, A., Olshtain, E., & Rosenstein, D. (1986). Advanced EFL apologies: What remains to be learned? *International Journal of the Sociology of Language, 62*, 51-74.

Cohen, A., & Upton, T. (2006). *Strategies in responding to the New TOEFL reading tasks.* (TOEFL Monograph Series Report No. 33). Princeton, NJ: ETS.

Cordón, L., & Day, J. (1996). Strategy use on standardized reading comprehension tests. *Journal of Educational Psychology, 88*(2), 288-295.

Creswell, J. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage Publications.

Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. (2005). *A teacher-verification study of speaking and writing prototype tasks for a new TOEFL* (TOEFL-MS-26). Princeton, NJ: ETS.

Cumming, A., Kantor, R., Baba, K., Eouanzoui, K., Erdosy, U., & James, M. (2006). *Analysis of discourse features and verification of scoring levels for independent and integrated prototype writing tasks for new TOEFL*. (TOEFL Monograph No. MS-30). Princeton, NJ: ETS.

Cunningham, A., Stanovich, K., & Wilson, M. (1990). Cognitive variation in adult college students differing in reading ability. In T. Carr & B. Levy (Eds.), *Reading and its development: Component skills approaches* (pp. 129-159). San Diego, CA: Academic Press.

Cziko, G. (1978). Differences in first- and second-language reading: The use of syntactic, semantic and discourse constraints. *The Canadian Modern Language Review*, *34*, 473-489.

Cziko, G. (1980). Language competence and reading strategies: A Comparison of first- and second-language oral reading errors. *Language Learning*, *30*, 101-116.

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (2002). *Dictionary of language testing*. Studies in Language Testing, 7. Cambridge, UK: Cambridge University Press.

Davies, A., & Widdowson, H. (1974). Reading and writing. In J. Allen & S. Corder (Eds.), *The Edinburgh course in applied linguistics, 3, Techniques in applied linguistics* (pp. 155-201). London, UK: Oxford University Press.

Desimone, L., & Le Floch, K. (2004). Are we asking the right questions? Using cognitive interviews to improve surveys in education research. *Educational Evaluation and Policy Analysis, 26*(1), 1-22.

Devine, J. (1987). General language competence and adult second language reading. In J. Devine, P. Carrell, & D. Eskey (Eds.), *Research on reading English as a second language* (pp. 73-86). Washington, DC: Teaching English to Speakers of Other Languages.

Devine, J. (1988). A case study of two readers: Models of reading and reading performance. In J. Devine, P. Carrell, & D. Eskey (Eds.), *Interactive approach to second language reading* (pp. 127-130). Cambridge, UK: Cambridge University Press.

Dole, J., Duffy, G., Roehler, L., & Pearson, P. (1991). Moving from the old to the new: Research in reading comprehension instruction. *Review of Educational Research, 61*, 239-264.

Dollerup, C., Glahn, E., & Rosenberg-Hansen, C. (1982). Reading strategies and test-solving techniques in an EFL reading comprehension test: A preliminary report. *Journal of Applied Language Study*, *1*(1), 93-99.

Douglas, D. (1981). An exploratory study of bilingual reading proficiency. In S. Hudelson (Ed.), *Learning to read in different languages* (pp. 33-102). Washington, DC: Center for Applied Linguistics.

Douglas, D. (2002). *Assessing languages for specific purposes*. Cambridge, UK: Cambridge University Press.

Drum, P., Calfee, R., & Cook, L. (1981). The effects of surface structure variables on performance in reading comprehension tests. *Reading Research Quarterly, 16,* 486-514.

Duran, R. (1989). Testing of linguistic minorities. In R. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 573-588). New York, NY: American Council on Education & Macmillan.

Ellis, R. (1994). *The study of second language acquisition*. Oxford, UK : Oxford University Press.

Ellis, R. (2001). *Form-focused instruction and second language learning*. Oxford, UK: Blackwell.

Enright, M., Grabe, W., Koda, K., Mosenthal, P., Mulcahy-Ernt, P., & Schedl, M. (2000). *TOEFL 2000 reading framework: A working paper.* (TOEFL Monograph No. MS-17). Princeton, NJ: ETS.

Enright, M., & Schedl, M. (2000). *Reading for a reason: Using reader purpose to guide test design*. Princeton, NJ: ETS.

Ericsson, K., & Simon, H. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: Massachusetts Institute of Technology Press.

ETS. (2002). *LanguEdge courseware score interpretation guide*. Princeton, NJ: Author.

ETS. (2003). *Task specifications for Next Generation TOEFL reading test.* Princeton, NJ: Author.

ETS. (2004). *English language competency descriptors*. Princeton, NJ: Author.

ETS. (2005). *Standard setting materials for the Internet-based TOEFL test*. Princeton, NJ: Author.

ETS. (2007). *TOEFL iBT tips: How to prepare for the TOEFL iBT*. Princeton, NJ: Author.

ETS. (2008). *TOEFL research—Ensuring test quality*. Princeton, NJ: Author.

ETS. (2009). *The official guide to the TOEFL test*. New York, NY: McGraw-Hill.

ETS. (2011a). *Reliability and comparability of TOEFL iBT scores*. Princeton, NJ: Author.

ETS. (2011b). *Validity evidence supporting the interpretation and use of TOEFL iBT scores*. Princeton, NJ: Author.

Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurement, 27*(3), 209-226.

Flick, U. (2009). *An introduction to qualitative research*. London, UK: Sage Publications.

Foltz, P. (1996). Comprehension, coherence, and strategies in hypertext and linear text. In J. Rouet, J. Levonen, A. Dillon, & R. Spiro (Eds.), *Hypertext and cognition* (pp. 109-136). Mahwah, NJ: Lawrence Erlbaum Associates.

Forster, D., Karn, R., Suzuki, S. & Tateyama, T. (1997). *Strategies 1: Building TOEIC/TOEFL test-taking skills.* Tokyo, Japan: Aratake Publishing Co.

Garner, R., & Krauss, K. (1982). Good and poor comprehender differences in knowing and regulating reading behaviors. *Educational Research Quarterly, 6*, 5-12.

Gass, S., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Goldman, S. (1997). Learning from text: Reflections on the past and suggestions for the future. *Discourse Processes, 23,* 357-398.

Gomez, G., Noah, A., Schedl, M., Wright, C., & Yolkut, A. (2007). Proficiency descriptors based on a scale-anchoring study of the new TOEFL iBT reading test. *Language Testing, 24*(3), 417-444.

Gordon, C. (1987). *The effect of testing method on achievement in reading comprehension tests in English as a foreign language* (Master's thesis). Tel Aviv, Israel: Tel Aviv University.

Grabe, W. (1986). The transition from theory to practice in teaching reading. In F. Dubin, D. Eskey, & W. Grabe (Eds.), *Teaching second language reading for academic purposes* (pp. 25-48). Reading, MA: Addison-Wesley.

Grabe, W. (1991). Current developments in second language reading research. *TESOL Quarterly, 25,* 375-406.

Green, A. (1998). *Verbal protocol analysis in language testing research: A handbook.* Studies in language testing, 5. Cambridge, UK: Cambridge University Press.

Green, J. (1987). In search of meaning: A sociolinguistic perspective on lesson construction and reading. In D. Bloome (Ed.), *Literacy and schooling* (pp. 4-34). Norwood, NJ: Ablex Publishing Corp.

Greene, S., & Higgins, L. (1994). "Once upon a time": The use of retrospective accounts in building theory in composition. In P. Smagorinsky (Ed.), *Speaking about writing* (pp. 115-140). Thousand Oaks, CA: Sage Publications.

Grotjahn, R. (1987). How to construct and evaluate a C-Test: A discussion of some problems and some statistical analyses. In R. Grotjahn, C. Klein-Braley, & D. Stevenson (Eds.), *Taking their measure: The validity and validation of language tests* (pp. 219-254). Bochum, Germany: Brockmeyer.

Gundel, J., & Tarone, E. (1983). Language transfer and the acquisition of pronominal anaphora. In S. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 281-296). Rowley, MA: Newbury House.

Guthrie, J. (1988). Locating information in documents: Examination of a cognitive model. *Reading Research Quarterly, 23*, 178-199.

Hauptman, P. (1979). A comparison of first and second language reading strategies among English-speaking university students. *Interlanguage Studies Bulletin*, *4*, 173-201.

Hawley, P. (2003). *Being bright is not enough: The unwritten rules of doctoral study* (2nd ed.). Springfield, IL: Charles C. Thomas Publishers.

Haynes, M., & Carr, T. (1990). Writing system background and second language reading: A component skills analysis of English reading by native-speaking readers of Chinese. In T. Carr & B. Levy (Eds.), *Reading and its development: Component skills approaches* (pp. 375-421). San Diego, CA: Academic Press.

Hill, C., & Parry, K. (1992). The test at the gate: Models of literacy in reading assessment. *TESOL Quarterly, 26,* 433-461.

Homburg, T., & Spaan, M. (1982). ESL reading proficiency assessment: Testing strategies. In M. Hines & W. Rutherford (Eds.), *On TESOL '81* (pp. 25-33). Washington, DC: Teaching English to Speakers of Other Languages.

Horiba, Y. (1993). Second language readers' memory for narrative texts: Evidence for structure-preserving top-down processing. *Language Learning, 43,* 345-372.

Hsiao, T-Y. (2004). Testing a social psychological model of strategy use with students of English as a foreign language. *Psychological Reports, 95,* 1059-1071.

Hudson, T. (1996). *Assessing second language academic reading from a communicative perspective: Relevance for TOEFL 2000* (TOEFL Monograph No. MS-4). Princeton, NJ: ETS.

Irwin, J. (1986). Cohesion and comprehension: A research review. In J. Irwin (Ed.), *Understanding and teaching cohesion comprehension* (pp. 31-34). Newark, DE: International Reading Association.

Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P., & Taylor, C. (2000). *TOEFL 2000 framework: A working paper.* (TOEFL Monograph No. MS-16). Princeton, NJ: ETS.

Jourdenais, R. (2001). Protocol analysis and SLA. In P. Robinson (Ed.), *Cognition and second language acquisition* (pp. 354-375). New York, NY: Cambridge University Press.

Just, M., & Carpenter, P. (1987). *The psychology of reading and language comprehension.* Boston, MA: Allyn & Bacon.

Kasper, G. (1998). Analyzing Verbal Protocols. *TESOL Quarterly*, *32*(2), 358-362.

Kelle, U. (2001). Sociological explanations between micro and macro and the integration of qualitative and quantitative methods. *Forum: Qualitative Social Research, 2*(1), 95-117.

Kelly, D. (2009). *Methods for evaluating interactive information retrieval systems with users*. Hanover, MA: Now Publishers.

Kirsch, I., & Mosenthal, P. (1990). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly, 25*, 5-30.

Kleiman, A., Cavalcanti, M., Terzi, S. & Ratto, I. (1986). *Perception of the lexicon and its discourse function: Some conditioning factors*. Campinas, Brazil: Universidade Estadual de Campinas.

Klein-Braley, C. (1985). A close-up on the C-Test: A study in the construct validation of authentic tests. *Language Testing, 2*, 76-104.

Klein-Braley, C. & Raatz, U. (1984). A survey of research on the C-Test. *Language Testing, 1*, 134-146.

Kobayashi, M. (2002). Method effects on reading comprehension test performance: Text organization and response format. *Language Testing, 19*(2), 193-220.

Koda, K., Takahashi, E., & Fender, M. (1998). Effects of L1 processing experience on L2 intra-word awareness. *Ilha do Desterro*: *A Journal of English Language, Literatures in English and Cultural Studies, 35*, 57-85.

Konecki, K. (2008). Triangulation and dealing with the realness of qualitative research. *Qualitative Sociology Review*, *4*(3), 7-28.

Kymes, A. (2008). Reconceptualizing reading strategy use in online environments. In K. Mokhtari & R. Sheorey (Eds.), *Reading strategies of first-and second-language learners: See how they read* (pp. 185-196). Norwood, MA: Christopher-Gordon.

Leighton, J. (2004). Avoiding misconception, misuse, and missed opportunities: The collection of verbal reports in educational achievement testing. *Educational Measurement: Issues and Practice, 23*(4), 6-15.

Liu, L., Schedl, M., Malloy, J., Kong, N., & ETS. (2009). *Does content knowledge affect TOEFL iBT reading performance? A confirmatory approach to differential item functioning*. (TOEFL iBT Research Report No. 09). Princeton, NJ: ETS.

Long, M., & Crookes, G. (1992). Three approaches to task-based syllabus design. *TESOL Quarterly, 26*, 27-56.

Lumley, T., & Brown, A. (2004). Test-taker and rater perspectives on integrated reading and writing tasks in the Next Generation TOEFL. *Language Testing Update, 35*, 75-79.

Lumley, T., & Brown, A. (2006). *Test taker response to integrated reading/writing tasks in TOEFL: Evidence from writers, texts, and raters*. Melbourne, Australia: Language Testing Research Centre, University of Melbourne.

Lunzer, E., Waite, M., & Dolan, T. (1979). Comprehension and comprehension skills. In E. Lunzer & K. Garnder (Eds.), *The effective use of reading* (pp. 37-71). London, UK: Heinemann.

Mackey, A., & Gass, S. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum Associates.

MacLean, M., & d'Anglejan, A. (1986). Rational cloze and retrospection: Insights into first and second language reading comprehension. *The Canadian Modern Language Review, 42*(4), 814-826.

MacWhinney, B., & Bates, E. (1989). *The crosslinguistic study of sentence processing*. Cambridge, UK: Cambridge University Press.

Mangubhai, F. (1990). Towards a taxonomy of strategies used by ESL readers of varying proficiencies while doing cloze exercises. *Australian Journal of Reading, 13*(2), 128-139.

Mangubhai, F. (1991). The processing behaviors of adult second language learners and their relationship to second language proficiency. *Applied Linguistics, 12,* 268-298.

Mann, S. (1982). Verbal reports as data: A focus on retrospection. In F. Katamba, S. Dingwall & S. Mann (Eds.), *Methods and problems in doing applied linguistic research* (pp. 87-104). Lancaster, UK: Department of Linguistics and Modern English Language, University of Lancaster.

McKeown, M., Beck, I., Sinatra, G., & Losterman, J. (1992). The contribution of prior knowledge and coherent text to comprehension. *Reading Research Quarterly, 27,* 79-93.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational Measurement (*3rd ed.*)* (pp. 13-103). New York, NY: American Council on Education & Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

Mislevy, R. (1993). Foundations of a new test theory. In N. Frederiksen, R. Mislevy, & I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 19-39). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mokhtari, K., & Reichard, C. (2008). The impact of reading purpose on the use of reading strategies. In K. Mokhtari & R. Sheorey (Eds.), *Reading strategies of first-and second-language learners: See how they read* (pp. 85-97). Norwood, MA: Christopher-Gordon.

Mokhtari, K., & Sheorey, R. (2008). *Reading strategies of first- and second-language learners: See how they read*. Norwood, MA: Christopher-Gordon.

Mokhtari, K., Sheorey, R., & Reichard, C. (2008). Measuring the reading strategies of first- and second-language readers. In K. Mokhtari & R. Sheorey (Eds.), *Reading strategies of first- and second-language learners: See how they read* (pp. 43-65). Norwood, MA: Christopher-Gordon.

Moran-Ellis, J., Alexander, V., Cronin, A., Dickinson, M., Fielding, J., Sleney, J., et al. (2006). Triangulation and integration: Processes, claims and implications. *Qualitative Research*, *6*(1), 45-59.

Muljani, M., Koda, K., & Moates, D. (1998). Development of L2 word recognition: A connectionist approach. *Applied Psycholinguistics, 19*, 99-114.

Munby, J. (1978). *Communicative syllabus design.* Cambridge, UK: Cambridge University Press.

Myers, M., & Paris, S. (1978). Children's metacognitive knowledge about reading. *Journal of Educational Psychology, 70,* 680-690.

Nevo, N. (1989). Test-taking strategies on a multiple-choice test of reading comprehension. *Language Testing*, *6*(2), 199-215.

Nikolov, M. (2006). Test-taking strategies of 12-13-year-old Hungarian learners of EFL: Why whales have migraine. *Language Learning, 57*(1), 1-51.

Oxford, R. (1990). *Language learning strategies: What every teacher should know.* Boston, MA: Heinle & Heinle.

Oxford, R. (1996). Employing a questionnaire to assess the use of language learning strategies. *Applied Language Learning, 7*(1-2), 25-45.

Patton, M. (2004). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage Publications.

Peirce, B. (1992). Demystifying the TOEFL reading test. *TESOL Quarterly, 26,* 665-691.

Perfetti, C. (1997). Sentences, individual differences, and multiple texts: Three issues in text comprehension. *Discourse Processes, 23,* 337-355.

Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing, 20*(1), 26-56.

Phakiti, A. (2006). Modeling cognitive and metacognitive strategies and their relationships to EFL reading test performance. *Melbourne Papers in Language Testing, 11,* 53-102.

Phakiti, A. (2008). Strategic competence as a fourth-order factor model: A structural equation modeling approach. *Language Assessment Quarterly, 5*(1), 20-42.

Polio, C., Gass, S., & Chapin, L. (2006). Using stimulated recall to investigate native speaker perceptions in native-nonnative speaker interaction. *Studies in Second Language Acquisition, 28*(2), 237-267.

Pollatsek, A., & Rayner, K. (1990). Lexical access: Some comments on models and metaphors. In D. Balota, G. d'Arcias, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 143-163). Hillsdale, NJ: Lawrence Erlbaum Associates.

Poole, A. & Mokhtari, K. (2008). ESL students' use of reading strategies when reading online and in print. In K. Mokhtari & R. Sheorey (Eds.), *Reading strategies of first-and second-language learners: See how they read* (pp. 197-213). Norwood, MA: Christopher-Gordon.

Powers, D. (1993). Coaching for the SAT: A summary of the summaries and an update. *Educational Measurement: Issues and Practice, 12,* 24-39.

Powers, D., & Wilson Leung, S. (1995). Answering the new SAT reading comprehension questions without the passages. *Journal of Educational Measurement*, *32*(2), 105-129.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Pressley, M., & Gaskins, I. (2006). Metacognitively competent reading comprehension is constructively responsive reading: How can such reading be developed in students? *Metacognition and Learning*, *1*(1), 99-113.

Purpura, J. (1997). An analysis of the relationships between test takers' cognitive and metacognitive strategy use and second language test performance. *Language Learning*, *47*(2), 289-325.

Purpura, J. (1998). Investigating the effects of strategy use and second language test performance with high- and low-ability test takers: A structural equation modeling approach. *Language Testing, 15*(3), 333-379.

Purpura, J. (1999). *Strategy use and second language test performance: A structural equation modeling approach.* Cambridge, UK: Cambridge University Press.

Robinson, M. (1992). Introspective methodology in interlanguage pragmatics research. In G. Kasper (Ed.), *Pragmatics of Japanese as native and target language* (pp. 27-82). Honolulu, HI: University of Hawaii Press.

Roeschl-Heils, A., Schneider, W., & van Kraayenoord, C. (2003). Reading literacy, metacognition and motivation: A follow-up study of German students in Grades 7 and 8. *European Journal of Psychology of Education, 18,* 75-86.

Rogers, W., & Bateson, D. (1991). The influence of test-wiseness on performance of high school seniors on school leaving examinations. *Applied Measurement in Education, 4*(2), 159-183.

Roller, C. (1990). Commentary: The interaction of knowledge and structure variables in the processing of expository prose. *Reading Research Quarterly, 25,* 79-89.

Rosenfeld, M., Leung, S., & Oltman, P. (2001). *The reading, writing, speaking, and listening tasks important for academic success at the undergraduate and graduate levels.* (TOEFL Monograph Series Rep. No. 21). Princeton, NJ: ETS.

Ross, S. (1997). Listener inference on a second language test. In G. Kasper & E. Kellerman (Eds.), *Advances in communication strategies research* (pp. 216-237). Harlow, UK: Longman.

Rupp, A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple choice questions shapes the construct: A cognitive processing perspective. *Language Testing, 23*(4), 441-474.

Rupp, A., Vock, M., Harsch, C., & Köller, O. (2008). *Developing standards-based assessment tasks for English as a first foreign language —Context, processes, and outcomes in Germany*. Münster, Germany: Waxmann.

Rutherford, W. (1983). Language typology and language transfer. In S. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 358-370). Rowley, MA: Newbury House.

Sanz, C. (2005). *Mind and context in adult second language acquisition: Methods, theory, and practice*. Washington, DC: Georgetown University Press.

Sawaki, Y., & Nissan, S. (2009). *Criterion-related validity of the TOEFL iBT listening section*. Princeton, NJ: ETS.

Sawaki, Y., Stricker, L., & Oranje, A. (2008). *Factor structure of the TOEFL Internet-based test (iBT): Exploration in a field trial sample.* (TOEFL iBT Research Report No.TOEFLiBT-04). Princeton, NJ: ETS.

Schedl, M., Gordon, A., Carey, P., & Tang, K. (1996). *An analysis of the dimensionality of TOEFL reading comprehension items*. (TOEFL Research Report No. 53). Princeton, NJ: ETS.

Sheorey, R., & Mokhtari, K. (2001). Differences in the metacognitive awareness of reading strategies among native and non-native readers. *System, 29*(4), 431-449.

Singhal, M. (2001). Reading proficiency, reading strategies, metacognitive awareness and L2 readers. *The Reading Matrix, 1*(1), 1-23.

Smith, S. (1979). Strategies, language transfer and the simulation of the second language learner's mental operations. *Language Learning, 29*, 345-361.

Stemmer, B. (1991). *What's on a C-test taker's mind: Mental processes in C-test taking.* Bochum, Germany: Brockmeyer.

Storey, P. (1995). *Developing and validating a taxonomy of test-taking strategies*. In Proceedings of the conference on Testing and Evaluation in Second Language Education, Hong Kong University of Science and Technology.

Storey, P. (1997). Examining the test-taking process: A cognitive perspective on the discourse cloze test. *Language Testing*, *14*(2), 214-231.

Stricker, L., & Rock, D. (2008). *Factor structure of the TOEFL Internet-based test across subgroups.* Princeton, NJ: ETS.

Symons, S., & Pressley, M. (1993). Prior knowledge affects text search success and extraction of information. *Reading Research Quarterly*, *28*, 251-259.

Taylor, C., & Angelis, P. (2008). The evolution of TOEFL. In C. Chapelle, M. Enright, & J. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 27-54). New York, NY: Routledge.

Taylor, C., Jamieson, J., Eignor, D., & Kirsch, I. (1998). *The relationship between computer familiarity and performance on computer-based TOEFL test tasks.*(TOEFL Research Report No. 61). Princeton, NJ: ETS.

Taylor, K., & Dionne, J-P. (2000). Accessing problem-solving strategy knowledge: The complementary use of concurrent verbal protocols and retrospective debriefing. *Journal of Educational Psychology*, *92*, 413-425.

Tian, S-P. (2000). *TOEFL reading comprehension: Strategies used by Taiwanese students with coaching-school training* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9976766)

Tsagari, C. (1994). *Method effect on testing reading comprehension: How far can we go?*(Master's thesis). Lancaster, UK: University of Lancaster.

Upton, T. (1997). First and second language use in reading comprehension strategies of Japanese ESL students. *TESL-EJ, 3*(1).

Upton, T. & Lee-Thompson, L. (2001).The role of the first language in second language reading. *Studies in Second Language Acquisition, 23*(4), 469-495.

Van Dijk, T. (1977). *Text and context.* London, UK: Longman.

Wadden, P., & Hilke, R. (1999). Comments on Liz Hamp-Lyons' "Ethical Test Preparation Practice: The Case of the TOEFL": Polemic Gone Astray: A Corrective to Recent Criticism of TOEFL Preparation. *TESOL Quarterly, 33*(2), 263-270.

Wall, D., & Horák, T. (2006). *The impact of changes in the TOEFL examination on teaching and learning in Central and Eastern Europe: Phase 1, The baseline study.* (TOEFL Monograph Series, MS-34). Princeton, NJ: ETS.

Wall, D., & Horák, T. (2008a). *The impact of changes in the TOEFL examination on teaching and learning in Central and Eastern Europe: Phase 2, Coping with change*. (TOEFL iBT Research Report, TOEFL iBT-05). Princeton, NJ: ETS.

Wall, D., & Horák, T. (2008b). *The impact of changes in the TOEFL examination on teaching and learning in a sample of countries in Europe: Phase 3, The role of the coursebook. Phase 4, Describing change*. Lancaster, UK: Department of Linguistics and English Language, Lancaster University.

Wallen, N., & Fraenkel, J. (2006). *How to design and evaluate research in education*. Boston, MA: McGraw-Hill.

Wang, L., Eignor, D., & Enright, M. (2008). A final analysis. In C. Chapelle, M. Enright, & J. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 259-318). New York, NY: Routledge.

Wijgh, I. (1995). A communicative test in analysis: Strategies in reading authentic texts. In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 154-170). Clevedon, UK: Multilingual Matters Ltd.

Xi, X. (2008). *Investigating the criterion-related validity of the TOEFL speaking scores for ITA screening and setting standards for ITAs*. (TOEFL iBT Research Report No.TOEFLiBT-03). Princeton, NJ: ETS.

Xi, X. (2010). How do we go about investigating test fairness? *Language Testing, 27*(2), 147-170.

Yamashita, J. (2003). Processes of taking a gap-filling test: Comparison of skilled and less skilled EFL readers. *Language Testing*, *20*(3), 267-293.

Yang, P. (2000). *Effects of test-wiseness upon performance on the Test of English as a Foreign Language* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database.(UMI No. NQ59700)

Yang, Y. (2002). Reassessing readers' comprehension monitoring. *Reading in a Foreign Language, 14*, 18-42.

Yien, L. (2001). Effective test-taking strategies on English tests: Implications from Taiwanese students. *Hong Kong Journal of Applied Linguistics, 6*(2), 22-43. (ERIC Document Reproduction Service No. EJ655676)

Yoshizawa, K. (2002). *Relationships among strategy use, foreign language aptitude, and second language proficiency: A structural equation modeling approach* (Doctoral

dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3040378)

yWorks GmbH.(2010). yEd (Version 3.6.1.1) [Computer software]. Available from yWorks' website: http://www.yWorks.com

Zhang, Y. (2008). *Repeater analyses for TOEFL iBT*. (ETS Research Memorandum No. RM. 08-05). Princeton, NJ: ETS.

APPENDICES

APPENDIX A

Propositions and Related Evidence in the TOEFL Validity Argument

Proposition	Evidence
The content of the test is relevant to and representative of the kinds of tasks and written and oral texts that students encounter in college and university settings.	Reviews of research and empirical studies of language use at English-medium institutions of higher education
Tasks and scoring criteria are appropriate for obtaining evidence of test takers' academic language abilities.	Pilot and field studies of task and test design; systematic development of rubrics for scoring written and spoken responses
Academic language proficiency is revealed by the linguistic knowledge, processes, and strategies test takers use to respond to test tasks.	Investigations of discourse characteristics of written and spoken responses and strategies used in answering reading comprehension questions
The structure of the test is consistent with theoretical views of the relationships among English language skills.	Factor analyses of a field-study test form
Performance on the test is related to other indicators or criteria of academic language proficiency.	Relationships between test scores and self-assessments, academic placements, local assessments of international teaching assistants, performance on simulated academic listening tasks
The test results are used appropriately and have positive consequences.	Development of materials to help test users prepare for the test and interpret test scores appropriately; long-term empirical study of test impact (washback)

Source: ETS (2008, p. 2).

APPENDIX B

Types of Texts and Tasks for the TOEFL 2000 Reading Test

<table>
<tr><td></td><td colspan="6">A. Types of texts defined by pragmatic and rhetorical features</td></tr>
<tr><td></td><td colspan="3">EXPOSITION</td><td colspan="2">ARGUMENTATION
PERSUASION
EVALUATION</td><td>HISTORICAL
NARRATION</td></tr>
<tr><td>B. Reader tasks</td><td>define/
describe/
elaborate/
illustrate</td><td>compare/
contrast/
classify</td><td>problem/
solution</td><td>explain/
justify</td><td>persuade</td><td>narrate</td></tr>
<tr><td>identify/
interpret</td><td colspan="6">Reader Purpose: Reading to Find Information and Reading for Basic Comprehension</td></tr>
<tr><td>summarize</td><td colspan="6" rowspan="3">Reader Purpose: Reading to Learn</td></tr>
<tr><td>define/
describe/
elaborate/
illustrate</td></tr>
<tr><td>compare/
contrast/
classify</td></tr>
<tr><td>problem/
solution</td><td colspan="6" rowspan="4">Reader Purpose: Reading to Integrate</td></tr>
<tr><td>explain/
justify</td></tr>
<tr><td>persuade</td></tr>
<tr><td>narrate</td></tr>
</table>

Source: Enright et al. (2000, p. 30).

APPENDIX C

Item Type Descriptions for the TOEFL-iBT Reading Test

ETS has circulated descriptions and specifications for the types of task items to be used as part of the TOEFL-iBT reading test (2002, pp. 48-50; 2003, pp. 4-17), which appear as follows:

I. Basic Comprehension

- *Vocabulary:* "measure[s] examinee's ability to comprehend the meanings of individual words and phrases as used in the context of the passage" (p. 4).

Example:

Passage: "In the animal world the task of moving about is fulfilled in many ways. For some animals locomotion is accomplished by changes in body shape . . ."

The word locomotion in the passage is closest in meaning to

A. evolution

B. movement*

C. survival

D. escape

- *Pronoun reference:* "measure[s] examinee's ability to identify relationships between pronouns and other anaphoric devices and their antecedents/postcedents within the passage" (p. 6).

Example:

Passage: "…Roots anchor the plant in one of two ways or sometimes by a combination of the two. The first is by occupying a large volume of shallow soil around the plant's base with a *fibrous root system*, one consisting of many thin, profusely

branched roots. Since these kinds of roots grow relatively close to the soil surface, they effectively control soil erosion. Grass roots are especially well suited to this purpose. Fibrous roots capture water as it begins to percolate into the ground and so must draw their mineral supplies from the surface soil before the nutrients are leached to lower levels . . ."

The phrase this purpose in the passage refers to

A. combining two root systems

B. feeding the plant

C. preventing soil erosion*

D. leaching nutrients

- *Sentence simplification:* "measure[s] examinee's ability to identify essential information as they process complex sentences in extended texts without getting lost in less important details and elaborations" (p. 8).

Example:

[Note: the tested sentence is highlighted in the passage below.]

Which of the following best expresses the essential information in the highlighted sentence? *Incorrect* answer choices change the meaning in important ways or leave out essential information.

A. Functional applied-art objects cannot vary much from the basic patterns determined by the laws of physics.*

B. The function of applied-art objects is determined by basic patterns in the laws of physics.

C. Since functional applied-art objects vary only within certain limits, arbitrary decisions cannot have determined their general form.

D. The general form of applied-art objects is limited by some arbitrary decision that is not determined by the laws of physics.

Passage context:

". . . Although we now tend to refer to the various crafts according to the materials used to construct them—clay, glass, wood, fiber, and metal—it was once common to think of crafts in terms of function, which led to their being known as "applied arts." Approaching crafts from the point of view of function, we can divide them into simple categories: containers, shelters, and supports. There is no way around the fact that containers, shelters, and supports must be functional. The applied arts are thus bound by the laws of physics, which pertain to both the materials used in their making and the substances and things to be contained, supported, and sheltered. These laws are universal in their application, regardless of cultural beliefs, geography, or climate. If a pot has no bottom or has large openings in its sides, it could hardly be considered a container in any traditional sense. Since the laws of physics, not some arbitrary decision, have determined the general form of applied-art objects, they follow basic patterns, so much so that functional forms can vary only within certain limits. Buildings without roofs, for example, are unusual because they depart from the norm. However, not all functional objects are exactly alike; that is why we recognize a Shang Dynasty vase as being different from on Inca vase. What varies is not the basic form but the incidental details that do not obstruct the object's primary function . . ."

- *Factual information:* "measure[s] examinees' ability to identify responses to questions about important factual information that is explicitly stated in a text. The examinees' task is to match the information requested in the item stem to the information in the text that answers the question" (p. 10).

Example:

Passage: ". . . Sculptures must, for example, be stable, which requires an understanding of the properties of mass, weight distribution, and stress. Paintings must have rigid stretchers so that the canvas will be taut, and the paint must not deteriorate, crack, or discolor. These are problems that must be overcome by the artist because they tend to intrude upon his or her conception of the work. For example, in the early Italian Renaissance, bronze statues of horses with a raised foreleg usually had a cannonball under that hoof. This was done because the cannonball was needed to support the weight of the leg. In other words, the demands of the laws of physics, not the sculptor's aesthetic intentions, placed the ball there. That this device was a necessary structural compromise is clear from the fact that the cannonball quickly disappeared when sculptors learned how to strengthen the internal structure of a statue with iron braces (iron being much stronger than bronze) . . ."

According to paragraph 2, sculptors from the Italian Renaissance stopped using cannonballs in bronze statues of horses because

A. they began using a material that made the statues weigh less

B. they found a way to strengthen the statues internally*

C. the aesthetic tastes of the public had changed over time

D. the cannonballs added too much weight to the statues

- *Negative fact:* "measure[s] examinees' ability to verify what information is true and what information is NOT true or not included in the passage based on information that is explicitly stated in the passage. The examinees' task is to locate the relevant information in the passage and verify that 3 of the 4 options are true and/or that one of them is false" (p. 12).

Example:

Passage: ". . . Although we now tend to refer to the various crafts according to the materials used to construct them—clay, glass, wood, fiber, and metal—it was once common to think of crafts in terms of function, which led to their being known as "applied arts." Approaching crafts from the point of view of function, we can divide them into simple categories: containers, shelters, and supports. There is no way around the fact that containers, shelters, and supports must be functional. The applied arts are thus bound by the laws of physics, which pertain to both the materials used in their making and the substances and things to be contained, supported, and sheltered. These laws are universal in their application, regardless of cultural beliefs, geography, or climate. If a pot has no bottom or has large openings in its sides, it could hardly be considered a container in any traditional sense . . ."

The passage discusses applied-art objects in relation to all of the following EXCEPT

A. the techniques used in their construction*

B. the ways they have been classified

C. their function

D. the universality of the laws that govern them

II. Inferencing

- *Basic inference:* "measure[s] examinees' ability to comprehend an argument or an idea that is strongly implied but not explicitly stated in the text. For example, if an effect is cited in the passage, an inference question might ask about its cause. If a comparison is made, an inference question might ask for the basis of the comparison. From an explicit description of a new phenomenon, examinees could be asked to infer the characteristics of the old phenomenon" (p. 25).

Example:

Passage: ". . . The nineteenth century brought with it a burst of new discoveries and inventions that revolutionized the candle industry and made lighting available to all. In the early to mid-nineteenth century, a process was developed to refine tallow (fat from animals) with alkali and sulfuric acid. The result was a product called stearin. Stearin is harder and burns longer than unrefined tallow. This breakthrough meant that it was possible to make tallow candles that would not produce the usual smoke and rancid odor. Stearins were also derived from palm oils, so vegetable waxes as well as animal fats could be used to make candles . . ."

Which of the following can be inferred from paragraph 1 about candles before the nineteenth century?

A. They did not smoke when they were burned.

B. They produced a pleasant odor as they burned.

C. They were not available to all. *

D. They contained sulfuric acid

- *Rhetorical purpose*: "measure[s] examinees' ability to identify the author's underlying rhetorical purpose in employing particular expository features in the passage and in ordering the exposition in a particular way. Correct responses require proficiency at inferring the nature of the link between specific features or exposition and the author's rhetorical purpose" (p. 27).

Example:

Passage: ". . . Sculptures must, for example, be stable, which requires an understanding of the properties of mass, weight distribution, and stress. Paintings must have rigid stretchers so that the canvas will be taut, and the paint must not deteriorate, crack, or discolor. These are problems that must be overcome by the artist because they tend to intrude upon his or her conception of the work. For example, in the early Italian Renaissance, bronze statues of horses with a raised foreleg usually had a cannonball under that hoof. This was done because the cannonball was needed to support the weight of the leg . . ."

Why does the author discuss the bronze statues of horses created by artists in the early Italian Renaissance?

A. To provide an example of a problem related to the laws of physics that an artist tries to overcome*

B. To argue that fine artists are unconcerned with the laws of physics

C. To contrast the relative sophistication of modern artists in solving problems related to the laws of physics

D. To note an exceptional piece of art constructed without the aid of technology

- *Insert text:* “measure[s] examinees’ ability to understand the lexical, grammatical, and logical links between successive sentences. Examinees are asked to determine where to insert a new sentence into a section of the reading passage that is displayed to them” (p. 31).

Example:

Look at the four squares [■] that indicate where the following sentence could be added to the passage.

> All three of them have strengths and weaknesses, but none adequately answers all of the questions the paintings present.

Where would the sentence best fit?

Example of how boxes would appear in the passage text:

Passage: “Scholars offer three related but different opinions about this puzzle. ■ One opinion is that the paintings were a record of the seasonal migrations made by herds. ■ Because some paintings were made directly over others, obliterating them, it is probable that a painting’s value ended with the migration it pictured. ■ Unfortunately, this explanation fails to explain the hidden locations, unless the migrations were celebrated with secret ceremonies. ■”

Click on a square [■] to add the sentence to the passage. To select a different location, click on a different square.

III. Reading to Learn

- *Prose summary:* “These items measure examinees’ ability to understand the major ideas and the relative importance of information in a text. Examinees are asked to select the major text ideas by distinguishing them from minor

ideas or ideas that are not in the text. . .The completed summary represents an able reader's mental framework of the text. The prose summary, therefore, should require examinees to identify information relevant to the major contrast(s), argument(s), etc. . . ." (p. 15).

Example:

[*Note:* Full text is necessary to determine main points and to eliminate incorrect options. The complete passage is not included here.]

An introductory sentence for a brief summary of the passage is provided below. Complete the summary by selecting the THREE answer choices that express important ideas in the passage. Some sentences do not belong in the summary because they express ideas that are not presented in the passage or are minor ideas in the passage. *This question is worth 2 points.*

> This passage discusses fundamental differences between applied-art objects and fine-art objects.
>
> - □
> - □
> - □

Answer choices

- The fine arts are only affected by the laws of physics because of the limitations of the materials that are used.
- Applied-art objects are bound by the laws of physics in two ways: by the materials used to make them, and the function they are to serve.

- Crafts are known as “applied arts” because it used to be common to think of them in terms of their function.
- In the fine arts, artists must work to overcome the limitations of their materials, but in the applied arts, artists work in concert with their materials.
- Making fine-art objects stable requires an understanding of the properties of mass, weight, distribution, and stress.
- In the twentieth century, artists working in the fine arts often treat materials in new ways whereas applied arts specialists continue to think of crafts in terms of function.

- *Schematic table*: “These items measure examinees’ ability to conceptualize and organize major ideas and other important information from across the text . . . The schematic table task reflects an able reader’s mental framework of the text. It should require examinees to identify and organize information relevant to the major contrast(s), argument(s), etc. . . . Examinees must both select the correct options and organize them correctly in the schematic table for the responses to be scored correct” (ETS, 2002, p. 48). Example (ETS, 2002, p. 50):

[Note: Full text is necessary to determine main points and to eliminate incorrect options. The complete passage is not included here.]

Complete the table below to summarize information about the two types of art discussed in the passage. Match the appropriate statements to the types of art with which they are associated. *This question is worth 3 points.*

Types of art	Statements
The Applied Arts	Select 3 ▪ ▪ ▪
The Fine Arts	Select 2 ▪ ▪

Statements

- An object’s purpose is primarily aesthetic.
- Objects serve a functional purpose.
- The incidental details of objects do not vary.
- Artists work to overcome the limitations of their materials.
- The basic form of objects varies little across cultures.
- Artists work in concert with their materials.
- An object’s place of origin is difficult to determine.

Drag your answer choices to the spaces where they belong. To review the passage, click on View Text.

APPENDIX D

Cohen and Upton's (2006) Rubrics of Test-taking Strategies on

the Reading Section of the TOEFL-iBT

Cohen and Upton (2006) developed coding rubrics of test-taking strategies that their 32 ESL respondents used on the TOEFL-iBT reading tasks. These coding rubrics comprised test-management strategies and test-wiseness strategies.

Test-Management Strategies Coding Rubric (T)

Strategy	Description
T1	Goes back to the question for clarification: Rereads the question.
T2	Goes back to the question for clarification: Paraphrases (or confirms) the question or task.
T3	Goes back to the question for clarification: Wrestles with the question intent.
T4	Reads the question and considers the options before going back to the passage/portion.
T5	Reads the question and then reads the passage/portion to look for clues to the answer, either before or while considering options.
T6	Predicts or produces own answer after reading the portion of the text referred to by the question.
T7	Predicts or produces own answer after reading the question and then looks at the options (before returning to text).
T8	Predicts or produces own answer after reading questions that require text insertion (I-it types).
T9	Considers the options and identifies an option with an unknown vocabulary.
T10	Considers the options and checks the vocabulary option in context.
T11	Considers the options and focuses on a familiar option.
T12	Considers the options and selects preliminary option(s) (lack of certainty indicated).
T13	Considers the options and defines the vocabulary option.
T14	Considers the options and paraphrases the meaning.
T15	Considers the options and drags and considers the new sentence in context (I-it).
T16	Considers the options and postpones consideration of the option.
T17	Considers the options and wrestles with the option meaning.
T18	Makes an educated guess (e.g., using background knowledge or extra-textual knowledge).
T19	Reconsiders or double checks the response.
T20	Looks at the vocabulary item and locates the item in context.
T21	Selects options through background knowledge.
T22	Selects options through vocabulary, sentence, paragraph, or passage overall meaning (depending on the item type).
T23	Selects options through elimination of other option(s) as unreasonable based on background knowledge.

(Table continues.)

T24	Selects options through elimination of other option(s) as unreasonable based on paragraph/overall passage meaning.
T25	Selects options through elimination of other option(s) as similar or overlapping and not as comprehensive.
T26	Selects options through their discourse structure.
T27	Discards option(s) based on background knowledge.
T28	Discards option(s) based on vocabulary, sentence, paragraph, or passage overall meaning as well as discourse structure.

Test-wiseness Strategies Coding Rubric (TW)

Strategy	Description
TW1	Uses the process of elimination (i.e., selecting an option even though it is not understood, out of a vague sense that the other options couldn't be correct).
TW2	Uses clues in other items to answer an item under consideration.
TW3	Selects the option because it appears to have a word or phrase from the passage in it—possibly a key word.

Source: Cohen and Upton (2006, pp. 36-37).

APPENDIX E

TOEFL-iBT Score Report Descriptors

Levels (score range)	Descriptors
High (22-30)	▪ have a very good command of academic vocabulary and grammatical structure; ▪ can understand and connect information, make appropriate inferences, and synthesize ideas, even when the text is conceptually dense and the language is complex; ▪ can recognize the expository organization of a text and the role that specific information serves within the larger text, even when the text is conceptually dense; and ▪ can abstract major ideas from a text, even when the text is conceptually dense and contains complex language.
Intermediate (15-21)	▪ have a good command of common academic vocabulary, but still have some difficulty with high-level vocabulary; ▪ have a very good understanding of grammatical structure; ▪ can understand and connect information, make appropriate inferences, and synthesize information in a range of texts, but have more difficulty when the vocabulary is high level and the text is conceptually dense; ▪ can recognize the expository organization of a text and the role that specific information serves within a larger text, but have some difficulty when these are not explicit or easy to infer from the text; and ▪ can abstract major ideas from a text, but have more difficulty doing so when the text is conceptually dense.
Low (0-14)	▪ have a command of basic academic vocabulary, but their understanding of less common vocabulary is inconsistent; ▪ have limited ability to understand and connect information, have difficulty recognizing paraphrases of text information, and often rely on particular words and phrases rather than a complete understanding of the text; ▪ have difficulty identifying the author's purpose, except when that purpose is explicitly stated in the text or easy to infer from the text; and ▪ can sometimes recognize major ideas from a text when the information is clearly presented, memorable, or illustrated by examples, but have difficulty doing so when the text is more demanding.

Source: Gomez et al. (2007, pp. 440-441).

APPENDIX F

Participant Characteristics

				Academic		Self-rating of English	
ID	Age	Gender	Nationality	Level	Major	Overall proficiency	Reading proficiency
1	18	M	Emirati	BS	Safety Engineering	Intermediate	Beginning
2	33	M	Egyptian	PhD	Geology	High intermediate	High intermediate
3	19	M	Omani	BS	Industrial Engineering	Intermediate	Intermediate
4	22	M	Saudi	BS	Electrical Engineering	High intermediate	Intermediate
5	28	M	Saudi	MS	Mechanical Engineering	High intermediate	Intermediate
6	21	M	Omani	BS	Electrical Engineering	High intermediate	Intermediate
7	26	M	Saudi	BS	Electrical Engineering	High intermediate	Intermediate
8	25	M	Saudi	BS	Electrical Engineering	Advanced	High intermediate
9	22	M	Saudi	BS	Industrial Engineering	High intermediate	Intermediate
10	23	M	Kuwaiti	BS	Management Information Systems	Intermediate	Intermediate
11	24	M	Omani	BS	Industrial Engineering	High intermediate	High intermediate
12	30	F	Iraqi	MA	English	Advanced	High intermediate
13	27	F	Egyptian	MA	English	Advanced	Advanced
14	23	M	Omani	BS	Industrial Engineering	Intermediate	Intermediate
15	24	M	Omani	BS	Industrial Engineering	High intermediate	Intermediate
16	21	M	Omani	BS	Architectural Engineering	Intermediate	Intermediate
17	20	M	Kuwaiti	BS	Architectural Engineering	High intermediate	Intermediate

Notes. F=female, M=male, BS=Bachelor of Science, MS=Master of Science, MA=Master of Arts. (Table continues)

Appendix F: Participant Characteristics (continued)

18	20	M	Kuwaiti	BS	Architectural Engineering	Intermediate	Intermediate
19	19	M	Omani	BS	Civil Engineering	Intermediate	Intermediate
20	23	M	Saudi	BS	Industrial Engineering	High intermediate	Intermediate
21	20	M	Saudi	BS	Industrial Engineering	High intermediate	Intermediate
22	37	M	Iraqi	MA	English	High intermediate	High intermediate
23	27	M	Iraqi	MA	English	Advanced	Advanced
24	21	M	Kuwaiti	BS	Mechanical Engineering	Intermediate	Intermediate
25	26	F	Iraqi	MS	Food Science	Intermediate	Intermediate

Notes. F=female, M=male, BS=Bachelor of Science, MS=Master of Science, MA=Master of Arts.

APPENDIX G

Participant TOEFL-iBT Records and Scores on the Reading Sets

	TOEFL iBT Records						Scores on Reading Sets		
ID	No. of times taken	Preparation mode	Length of preparation in months	Year last taken	Total score /120	Reading score /30	Set 1 /14	Set 2 /14	Total /28
1	2	3	0	2010	56	15	9	9	18
2	1	1	3-4	2008	92	28	11	9	20
3	3	1	3-4	2009	81	23	7	10	17
4	2	2	1-2	2007	48	18	7	7	14
5	3	1	5-6	2010	72	15	11	11	22
6	3	1	1-2	2008	50	9	8	6	14
7	1	3	0	2006	61	9	7	8	15
8	5	1	7-8	2006	52	15	6	4	10
9	2	1	1-2	2009	38	9	2	2	4
10	5	1	11-12	2009	70	13	9	11	20
11	1	1	1-2	2007	92	22	9	10	19
12	4	1,2	1-2	2010	107	24	12	9	21
13	5	1	9-10	2009	103	25	10	14	24
14	3	1	1-2	2007	64	5	5	5	10
15	1	1	1-2	2005	65	13	11	14	25
16	5	1	3-4	2009	69	14	4	4	8

Notes. Preparation mode (1=self-preparation, 2=enrolled in a preparation program, 3=did not prepare). (Table continues)

Appendix G: Participant TOEFL-iBT records and scores on the reading sets (continued)

17	1	3	0	2009	96	18	9	9	18
18	1	1	1-2	2009	62	17	8	5	13
19	5	1	5-6	2009	62	10	7	9	16
20	6	1	1-2	2007	66	5	7	6	13
21	1	1	1-2	2008	45	3	8	4	12
22	2	1	1-2	2008	76	16	13	9	22
23	1	1	1-2	2008	74	20	9	11	20
24	2	3	0	2010	74	14	8	7	15
25	3	1	1-2	2010	61	14	4	8	12

Notes. Preparation mode (1=self-preparation, 2=enrolled in a preparation program, 3=did not prepare).

APPENDIX H

The IRB Approval of Research and Related Documents

Oklahoma State University Institutional Review Board

Date:	Tuesday, May 11, 2010
IRB Application No	AS1046
Proposal Title:	Test-Taking Strategy Use on the Reading Section of the TOEFL iBT: A Study of Arab ESL Learners
Reviewed and Processed as:	Expedited

Status Recommended by Reviewer(s): Approved Protocol Expires: 5/10/2011

Principal Investigator(s):

Mohammed Assiri
2213 N. Manning St.
Stillwater, OK 74075

Gene Halleck
205 Morrill
Stillwater, OK 74078

The IRB application referenced above has been approved. It is the judgment of the reviewers that the rights and welfare of individuals who may be asked to participate in this study will be respected, and that the research will be conducted in a manner consistent with the IRB requirements as outlined in section 45 CFR 46.

☒ The final versions of any printed recruitment, consent and assent documents bearing the IRB approval stamp are attached to this letter. These are the versions that must be used during the study.

As Principal Investigator, it is your responsibility to do the following:

1. Conduct this study exactly as it has been approved. Any modifications to the research protocol must be submitted with the appropriate signatures for IRB approval.
2. Submit a request for continuation if the study extends beyond the approval period of one calendar year. This continuation must receive IRB review and approval before the research can continue.
3. Report any adverse events to the IRB Chair promptly. Adverse events are those which are unanticipated and impact the subjects during the course of this research; and
4. Notify the IRB office in writing when your research project is complete.

Please note that approved protocols are subject to monitoring by the IRB and that the IRB office has the authority to inspect research records associated with this protocol at any time. If you have questions about the IRB procedures or need any assistance from the Board, please contact Beth McTernan in 219 Cordell North (phone: 405-744-5700, beth.mcternan@okstate.edu).

Sincerely,

Shelia Kennison, Chair
Institutional Review Board

Information for Participant Recruitment

Principally, contacted persons among Arab ESL learners will be offered participation in this research in Arabic, either in person or by phone. The English script of the information to be provided to each contacted person is as follows:

I, Mohammed Assiri, am going to conduct a study which aims to explore what Arab ESL learners do when responding to the reading section of the TOEFL iBT. If you are initially willing to participate in this study, I would like to ask you the following questions: First, have you ever taken the TOEFL iBT or at least prepared to take it? Second, when was the last time you took it? Third, have you ever dealt with the following topics when responding to the reading section of the TOEFL iBT (*Nineteenth-century Politics in the United States* and *Geology and Landscape*)?

[Note to the reader: The contacted person will be considered eligible to participate in this study if he or she either took the TOEFL iBT or prepared to take it before 2009 (the year in which the two reading sets to be used in this research were published for test preparation) or they assure the researcher that they have never responded or dealt with any of the topics of the reading sets (i.e., Reading set#1: *Nineteenth-century Politics in the United States* and Reading set#2: *Geology and Landscape*). Otherwise, the contacted person will be thanked for expressing his or her willingness to participate, and the reason why he or she was not considered eligible to participate will be clarified. The reason for this approach is to ensure that none of the participants is already familiar with any of the reading sets; otherwise, he or she will be advantaged in contrast to the other participants]. If the contacted person is eligible to participate in this research, he or she will be informed about the research as follows:

In this study, you will be requested to respond to two TOEFL-iBT reading sets, each set is basically a computerized practice test, each has a reading passage followed by 12 to 13 question items. While you are responding to each reading set, a screen recorder will be recording your activity on the computer screen. First, you will be asked to respond to the first reading set in 20 minutes. Then, you will be interviewed and asked how you responded to the first reading set. The screen recording will play as a movie to help you recall your thoughts during your response to the reading set by using the cursor motion in the playback. The researcher may ask you to describe, explain, or clarify your thoughts and/or response behaviors during the interview. The interview will be audio-recorded digitally throughout. Next, you will be offered to rest in between two data collection sessions. Afterwards, you will respond to the second reading set. The same procedures as those followed in the first data collection session will be followed in this session. At the end of the whole data collection procedure, you will be provided with your scores on the reading sets and rewarded for your participation a sum of money that amounts from $10 to $20 depending on your total score on the reading sets.

At the most, all procedures involved should take from 100 to 120 minutes to complete. The personal information you will be requested to provide is limited to your age, Arab nationality, your academic level (i.e., undergraduate or graduate) and major, the number of times you have taken the TOEFL iBT, how you prepared for the test and how long this preparation took, the last time you took the test and your TOEFL-iBT total score and score on the reading section, your evaluation of your levels of reading and overall proficiency in English, audio-recordings of the interview, and screen recordings of your response to the reading sets. Please rest assured that all your personal information

and screen- and audio-recordings as well as your scores on the reading sets, responses to the interviews, and verbal reports will be kept private and confidential so long as the researcher wishes to use the data for research purposes.

If you have any questions about the research, please feel free to contact:

Researcher: Mohammed Assiri. Phone: 405-762-1552. Email: massiri@okstate.edu

Research advisor: Dr. Gene Halleck. Phone: 405-744-9474. Email: gene.halleck@okstate.edu

If you have questions about your rights as a research volunteer, you may contact:

Chair of Institutional Review Board: Dr. Shelia Kennison. Phone: 405-744-3377. Email: irb@okstate.edu.

Arabic version:

أنا، محمد عسيري، أنوي أجراء دراسة تهدف لإستكشاف ما يقوم به الطلاب العرب الذين يتعاملون مع اللغة الإنجليزية كلغة ثانية عند تجاوبهم مع جزء القراءة من الـ TOEFL iBT . إذا كنت مبدئياً ترغب في المشاركة، بودي أن أسألك الثلاث الأسئلة التالية: أولا، هل سبق لك أن أخذت إختبار الـ TOEFL iBT أو التحضير لأخذه؟ ثانياً، متى كانت آخر مرة أخذت فيها هذا الإختبار؟ ثالثاً، هل سبق وتعاملت مع أي من الموضوعين التاليين عندما تجاوبت مع جزء القراءة من الـ TOEFL iBT:

Nineteenth-century Politics in the United States أو *Geology and Landscape*؟

في هذه الدراسة سيطلب منك الإجابة على مجموعتي قراءة. تأخذ كل مجموعة شكل إختبار على الكمبيوتر وتتألف من قطعة قراءة متبوعة بـ 12 إلى 13 سؤال على القطعة. أثناء تجاوبك مع مجموعة القراءة سيقوم مسجل مرئي بتسجيل حركات المؤشر وبالتالي نشاطك الكامل على شاشة الكمبيوتر. سيطلب منك التجاوب مع كل مجموعة قراءة في ظرف 20 دقيقة. بعد إنتهائك من كل مجموعة قراءة سيتم إجراء مقابلة معك حول كيفية تجاوبك معها. سيتم تشغيل التسجيل المرئي في شكل video حتى يساعدك على تذكر أدائك و كيفية تجاوبك مع مجموعة القراءة. قد

يسألك الباحث كي تصف، تعلل، أو توضح سلوكك التجاوبي مع قطعة القراءة. سيتم تسجيل المقابلة صوتياً. سيعرض عليك أخذ قسط من الراحة بين جلستين أو مقابلتين. البيانات التي ستطلب منك ستقتصر على العمر، الجنس، التخصص والدرجة العلمية التي تدرس لها، عدد المرات التي أخذت فيها إختبار الـ TOEFL iBT، كيفية إعدادك للإختبار وكم كان طول فترة الإعداد، ومتى كانت آخر مرة أخذت فيها الإختبار والدرجة التي حصلت عليها في الإختبار بشكل عام وفي جزء القراءة بشكل خاص، وتقييمك لكفاءتك اللغوية الإنجليزية بشكل عام وكفاءتك في القراءة بشكل خاص. البيانات الأخرى تتضمن التسجيل المرئي لأدائك لمجموعتي القراءة والصوتي لإخبارك اللفظي والمقابلة. جميع هذه الإجراءات ستجرى في الزمان والمكان المناسبين للمشارك شريطة أن يكون ذلك محفز للحصول على بيانات ذات جودة عالية (وقت مناسب وجو هادئ). تستغرق الإجراءات من 100 إلى 120 دقيقة. في نهاية الإجراءات ستزود بالنتائج التي حصلت عليها من خلال تعاملك مع مجموعتي القراءة وستكافأ مقابل مشاركتك مقابل مالي يتراوح من 10$ إلى 20$ تبعا للدرجة الكلية التي حصلت عليها من خلال تجاوبك مع مجموعتي القراءة.

الرجاء أن تثق تماماً أن كل البيانات الشخصية التي ستزود بها الباحث والتسجيلات الصوتية والمرئية وإجاباتك على المقابلة ونتائجك والإخبار اللفظي ستعامل بكامل الخصوصية والسرية ما دام الباحث بحاجة لهذه البيانات لأغراض بحثية.

يمكن للمشارك أو المشاركة الإتصال عند الرغبة في السؤال أو الإستفسار عن البحث بـ:

الباحث: محمد عسيري تلفون: 4057621552 إيميل: massiri@okstate.edu

الأستاذة المشرفة: د. جين هالك تلفون: 4057449474 إيميل:gene.halleck@okstate.edu

إذا أردت السؤال أو الإستفسار عن حقوقك كمشارك أو مشاركة في هذه الدراسة يمكنك الإتصال بـ:

رئيسة الجهة المعنية بحماية المشاركين وحفظ حقوقهم (Institutional Review Board)

د. شيليا كنيسين تلفون: 4057443377 إيميل: irb@okstate.edu

Informed Consent Document

Project Title: Test-taking strategy use on the reading section of the TOEFL iBT: A study of Arab ESL learners

Investigator: Mohammed Assiri, MA.

Purpose:

This research aims to look into aspects of test-taking strategy choice and use among Arab ESL learners when they respond to the reading tasks of the TOEFL internet-based test (iBT). It has three goals: To explore the kinds of test-taking strategies that selected respondents from the target population tend to use when responding to the TOEFL-iBT reading tasks, to determine if there are differences between high- and low-scoring test-takers among respondents in their choice and use of test-taking strategies on the reading tasks, and to find out what aspects of effective test-taking strategy use are exhibited by respondents with which reading tasks. As such, this research calls for human participation that involves each subject's response to two reading sets similar in format to actual TOEFL-iBT reading sets and verbal report of his or her response behaviors on the reading tasks.

Procedures:

First, you will be asked to respond to a participant background questionnaire. Then, you will be trained in how to produce verbal report of thoughts that you had in your mind while responding to a reading task. After that, you will respond to two reading sets, each is basically a computerized-practice test, each has a reading passage and 12 to 13 question items on the passage. While you are

responding to each reading set, a screen recorder will be recording your activity on the computer screen. You will be asked to respond to the first reading set within 20 minutes. Then, you will be interviewed and asked how you responded to the first reading set. The screen recording will play as a movie to help you recall your thoughts during your response to the reading set. The researcher may ask you to describe, explain, or clarify your thoughts and/or response behaviors during the interview. The interview will be audio-recorded digitally throughout. Next, you will be offered to rest in between two sessions. Afterwards, you will respond to the second reading set. The same procedures as those followed in the first session will be followed in this session.

You will only be requested to provide as personal information: your age, gender, academic level and major for which you had to take the test, the number of times you have taken the TOEFL iBT, how you prepared for the test and how long this preparation took, when the last time you took the test was and what your TOEFL-iBT total and reading scores were, and your evaluation of your level of reading and overall proficiency in English. Other personal data include audio- and screen- recordings. All the procedures will take place at the time and place of convenience to the participant provided that these conditions are conducive to optimal data collection (i.e., appropriate time and quiet atmosphere). Preferred places include public and university libraries. The procedures altogether are not expected to take more than 120 minutes.

Risks of Participation:

There are no known risks associated with this project which are greater than those ordinarily encountered in daily life.

Benefits:

There are no expected benefits from participating in this study except those of serving academic research. The knowledge drawn from this study is expected to enable specialists in language testing to determine what aspects of effective test-taking strategy use can be applied to which reading tasks and items on the TOEFL-iBT reading section, which will necessarily inform the practice of test-taking strategy training and instruction. There is also research evidence that participants engaged in this kind of research can benefit from having them bring to their consciousness or awareness those strategies that they tend to use, which enables them to judge the effectiveness of these strategies, especially the kinds of strategies used on academic reading tasks.

Confidentiality:

Please rest assured that all the records with your data will be kept private. Any written results will discuss group findings and will not include information that will identify you. Research records will be stored securely and only the researcher, the research advisor, and individuals responsible for research oversight will have access to the records. The data will be reported without referring to any personal information identifying you. During the data collection and analysis stages, numbers will be used to distinguish subjects from one another. In the presentation of the results and findings of this research,

pseudonyms or anonyms will be used to refer to individual cases among participants. The responses you will provide will not be used for other purposes than those serving academic research. Your responses, audio- and screen-recordings, and personal information will be kept private and confidential throughout the research stages. All of the data will be stored in the researcher's personal folders and external PC memory out of sight from anyone else. After the research is completed, the researcher will continue maintaining with total care the privacy and confidentiality of all data collected in order to use them in future research. The same precautions will be made in any future research to ensure full security and confidentiality of the data.

Compensation:

You will be rewarded a sum of money amounting from $10 to $20 depending on your total score on the reading sets that you will respond to. This amount of money includes $10 allotted to you in compensation for your participation in this study. The purpose of the reward is to motivate you to perform the reading tasks to the best of your ability and to try to achieve as high scores as you can, and make this experience of consequential value to you. This is anticipated to make you feel that the extent to which you perform the reading tasks will constitute a rewarding experience. This, in turn, will best serve the purpose of this research through the collection of strategy data that approximate the kind of data that would be collected in an actual test-taking experience in terms of quality.

Contacts:

Researcher: Mohammed Assiri. Phone: 405-762-1552.

Email: massiri@okstate.edu

Research advisor: Dr. Gene Halleck. Phone: 405-744-9474.

Email: gene.halleck@okstate.edu

If you have questions about your rights as a research volunteer, you may contact:

Chair of Institutional Review Board: Dr. Shelia Kennison. Phone: 405-744-3377.

Email: irb@okstate.edu.

Participant Rights:

Please note that your participation in this research is voluntary and that you can discontinue it at any time without reprisal or penalty.

Signatures:

I have read and fully understood the consent form. I sign it freely and voluntarily. A copy of this form has been given to me.

_____ _____

Signature of Participant Date

I certify that I have personally explained this document before requesting that the participant sign it.

_____ _____

Signature of Researcher Date

نموذج الموافقة على المشاركة في دراسة

موضوع البحث: إستخدام إستراتيجيات الإجابة عن جزء القراءة من التوفل (المدار على النت): دراسة للطلاب العرب الذين يتعاملون مع اللغة الإنجليزية كلغة ثانية.

الباحث: محمد عسيري، حاصل على مؤهل الماجستير في تدريس اللغة الإنجليزية كلغة ثانية.

الغرض من الدراسة: تهدف الدراسة إلى التعرف على كيفية تعامل الطلاب العرب الذين يتعاملون مع اللغة الإنجليزية كلغة ثانية مع جزء القراءة من التوفل (المدار على النت). للدراسة ثلاثة أهداف أساسية: إستكشاف الطرق التي يستخدمها أفراد عينة الدراسة عند الإجابة عن المهام القراءية، تحديد ما إذا كان هناك فروق بين الحاصلين على درجات عليا و الحاصلين على درجات دنيا في إختيار وإستخدام هذه الطرق، و التعرف على مظاهر الإستخدام الفعّال لهذه الطرق مع أي من المهام القراءية. وبالتالي هذا البحث يتطلب المشاركة الإنسانية التي تتضمن التجاوب مع مجموعتي قراءة مماثلة لتلك التي يحويها الإمتحان الفعلي والمراقبة الذاتية والإخبار اللفظي والمقابلة البعدية التابعة للتجاوب مع مجموعتي قراءة.

الإجراءات: أولا، سيطلب منك تعبئة إستبانة المشاركة. ثم سيتم تدريبك على كيفية الإخبار اللفظي وسرد الأفكار التي صاحبت أداءك لمهمة قراءية معينة. بعد ذلك سيطلب منك الإجابة على مجموعتي قراءة على شكل مبرمج على الكمبيوتر. كل مجموعة تحوي قطعة متبوعة بـ 12 إلى 13 سؤال على القطعة. أثناء تجاوبك مع مجموعة القراءة سيقوم مسجل مرئي بتسجيل تحركات المؤشر وبالتالي نشاطك الكامل على شاشة الكمبيوتر. سيطلب منك التجاوب مع كل مجموعة قراءة في ظرف 20 دقيقة. بعد إنتهائك من كل مجموعة قراءة سيتم إجراء مقابلة معك حول كيفية تجاوبك معها. سيتم تشغيل التسجيل المرئي في شكل video حتى يساعدك على تذكر أدائك و كيفية تجاوبك مع مجموعة القراءة. قد يسألك الباحث كي تصف، تعلل، أو توضح سلوكك التجاوبي مع قطعة القراءة. سيتم تسجيل المقابلة صوتياً. سيعرض عليك أخذ قسط من الراحة بين الجلستين. البيانات التي ستطلب منك ستقتصر على العمر، الجنس، التخصص والدرجة العلمية المطلوب لها الإختبار، عدد المرات التي أخذت فيها الـ TOEFL-iBT ، كيفية إعدادك للإختبار وكم كان طول فترة الإعداد، ومتى كانت آخر مرة أخذت فيها الإختبار والدرجة التي حصلت عليها آخر مرة في الإختبار بشكل عام وفي جزء القراءة بشكل خاص، وتقييمك لكفاءتك اللغوية الإنجليزية بشكل عام وكفاءتك في القراءة بشكل خاص. البيانات الأخرى تتضمن التسجيل الصوتي والمرئي لأدائك لمجموعتي القراءة.

جميع هذه الإجراءات ستجرى في الزمان والمكان المناسبين للمشارك شريطة أن يكون ذلك محفز للحصول على بيانات ذات جودة عالية (وقت مناسب وجو هادئ). لن تستغرق الإجراءات أكثر من 120 دقيقة.

مخاطر المشاركة: لا توجد هناك أي مخاطر مترتبة على المشاركة في هذه الدراسة.

عوائد المشاركة: لا توجد هناك أي عوائد مترتبة على المشاركة في هذه الدراسة عدا تلك التي تصب في مصلحة البحث الأكاديمي. المعرفة المستقاة من هذا البحث ستمكن المتخصصين في مجال التقييم اللغوي من تحديد مظاهر الإستخدام الفعال لطرق أخذ الإمتحانات المستخدمة مع أي من المهام القراءية في جزء القراءة من الـ -TOEFL iBT. وذلك سيساعد بالضرورة في تدعيم وتطوير برامج التدريب على إستراتيجيات أخذ الإمتحانات. هناك أبحاث تدل على أن المشاركين في هذا النوع من الإبحاث يستفيدون من عملية إستذكار الطرق التي يستخدمونها وجعلها في مجال الوعي أو الإدراك مما يساعدهم في تقييم هذه الإستراتجيات ولاسيما تلك التي تستخدم مع مهام القراءة الإكاديمية.

الخصوصية: ثق تمام الثقة أنه سيتم حفظ البيانات في ملفات الباحث الشخصية بعيدا عن نظر أي شخص أخر.لن يطلع على البيانات خلال فترة تحليلها وصياغة النتائج أحد سوى الباحث والأستاذة المشرفة على البحث أو أشخاص معنيين بحماية حقوق المشاركين في الأبحاث الإنسانية. لن تستخدم أي من البيانات لغير أغراض خدمة البحث العلمي. نتائج الدراسة ستناقش الإكتشافات التي سيتم التوصل إليها دون الكشف عن هوية المشاركين في الدراسة. خلال مرحلة جمع البيانات وتحليلها ستستخدم الأرقام للتعبير عن المشاركين. وخلال مرحلة تقديم النتائج والإستكشافات ستستخدم أسماء مستعارة للتعبير عن حالات فردية من المشاركين. ممن الممكن أن تخضع إجراءات الموافقة على المشاركة في الدراسة وجمع البيانات لمراقبة الأشخاص المعنيين بحماية حقوق المشاركين في الأبحاث. ستعامل جميع البيانات بغاية السرية والخصوصية لإستخدامها في أبحاث مستقبلية.

التعويض عن المشاركة: ستتم مكافئتك بمبلغ مالي يتراوح من 10$ إلى 20$ بحسب مجموع الدرجات الذي ستتحصل عليه من تجاوبك مع مجموعتي القراءة. يتضمن هذا المبلغ 10$ كمبلغ ممنوح لكل مشارك ومشاركة كتعويض مقابل المشاركة في هذه الدراسة. الهدف من المكافئة أو المبلغ الزائد هو تحفيزك بحيث تؤدي المهام القراءية بأحسن ما يمكنك وتحاول الحصول على أعلى درجة ممكنة، وتكون هذه التجربة ذات مردود أيجابي بالنسبة

لك. وهذا بدوره سيخدم الغرض من البحث من خلال جمع بيانات لاتختلف كثيراً عن تلك التي ممكن أن تجمع من خلال تجربة واقعية مع الإختبار من حيث الجودة.

التواصل: يمكن للمشارك أو المشاركة الإتصال عند الرغبة في السؤال أو الإستفسار عن البحث بـ:

الباحث: محمد عسيري تلفون: 4057621552 إيميل: massiri@okstate.edu

الأستاذة المشرفة: د. جين هالك تلفون:4057449474 إيميل: gene.halleck@okstate.edu

إذا أردت السؤال أو الإستفسار عن حقوقك كمشارك أو مشاركة في هذه الدراسة يمكنك الإتصال بـ:

رئيسة الجهة المعنية بحماية المشاركين وحفظ حقوقهم (Institutional Review Board)

د. شيليا كنيسين تلفون: 4057443377 إيميل: irb@okstate.edu

حقوق المشارك أو المشاركة: المشاركة في هذه الدراسة إختيارية ومن حق المشارك أو المشاركة عدم إكمال الدراسة إو الإنسحاب منها متى ما أراد أو أرادت دون أن يترتب على ذلك أي إشكال للمشارك أو المشاركة.

التوقيع

قرأت و فهمت نص نموذج الموافقة على المشاركة في الدراسة وعلى ذلك أوقع عن إختيار مني وتطوع للمشاركة. وقد أعطيت لي نسخة من هذا النموذج.

---------- ----------

توقيع المشارك أو المشاركة التاريخ

أؤكد على أني وضحت و شرحت نص هذا النموذج شخصيا قبل الطلب من المشارك أو المشاركة التوقيع.

---------- ----------

توقيع الباحث التاريخ

APPENDIX I

Participant Background Questionnaire

Participant ID#

Age: _____

Gender (circle)**:** female male

Native country: _____

Academic major: _____

What was the degree for which you had to obtain TOEFL-iBT scores (circle)**:**

1. bachelor of arts or sciences 2. master of arts or sciences 3. doctorate

How many times have you taken the TOEFL iBT? (Circle)**:**

1 2 3 4 5 6 7 8 9 10

If you have taken the TOEFL iBT more than 10 times, please give a number: ____

How did you prepare for the TOEFL iBT (circle)**:**

1. self-preparation 2. enrolled in a preparation program 3. did not prepare

If you prepared to take the TOEFL iBT, how long did this preparation take? Circle the correct choice!

1-2 months 3-4 months 5-6 months 7-8 months 9-10 months 11-12 months

If it took you longer than 12 months to prepare for the TOEFL iBT, please give a number: ____

In which year did you take the TOEFL iBT the last time? Circle the year!

2010 2009 2008 2007 2006 2005

What was your last TOEFL-iBT total score? _____ /120

What was your score on the TOEFL-iBT reading section the last time? ___ /30

How do you rate your overall proficiency in English? (Put a check mark ✓)

___beginning ___intermediate ___high intermediate ___advanced

How do you rate your reading proficiency in English? (Put a check mark ✓)

___beginning ___intermediate ___high intermediate ___advanced

How do you evaluate your level of motivation to perform the TOEFL-iBT reading tasks that are part of this research? (Put a check mark ✓)

___low ___below average ____average ___above average ___high

Participant Contact Information

Name (optional): _____

Phone number: _____

Email: _____

Please provide the researcher with both of your phone number and email address, or at least one of them. This will enable the researcher to contact you in case he wants some more clarification or explanation from you.

Thank you,

The researcher

APPENDIX J

Verbal Report Training

I. Introduction

Dear Participant:

After you respond to each reading set, I would like you to express your thinking or the thoughts that you had in your mind while answering each question item. You say aloud whatever you thought about while you were trying to answer each question item on the reading set. Please try to include every single thought and detail.

I would like to know how you dealt with the reading sets, how you thought while answering the questions, and how you chose the options that you thought are right and why.

I would like you to talk in Arabic; you can also use a mix of both Arabic and English.

Please note that

- your talk will be audio-recorded digitally.
- I may provide you with prompts and ask questions to know more about your response behaviors.

II. Modeling

Now, I will show you how to produce verbal report using a grammar exercise.

Directions: Which one of the sentences below would be the most correct answer to Speaker A's question?

Speaker A: Where did Ahmad go yesterday?

Speaker B: ……. **.**

□ He goed to see his grandma.

□ He went to see his grandma.

□ He has gone to see his grandma.

□ He goes to see his grandma.

III. Practice

Now, you will do some practice producing verbal report with one reading paragraph that has one question on it (or a reading mini-test).

After you pick the answer, please remember:

- to describe whatever you were thinking about while you were answering the question.
- to use Arabic; or a mix of both Arabic and English.
- that I may provide you with prompts and ask questions to know more about your response behaviors.

Please respond to the following reading and the question on it and then describe your response (**Time**: 2 minutes).

Paragraph: "…Sculptures must, for example, be stable, which requires an understanding of the properties of mass, weight distribution, and stress. Paintings must have rigid stretchers so that the canvas will be taut, and the paint must not deteriorate, crack, or discolor. These are problems that must be overcome by the artist because they tend to intrude upon his or her conception of the work. For example, in the early Italian Renaissance, bronze statues of horses with a raised foreleg usually had a cannonball under that hoof. This was done because the cannonball was needed to support the weight of the leg. In other words, the demands of the laws of physics, not the sculptor's aesthetic

intentions, placed the ball there. That this device was a necessary structural compromise is clear from the fact that the cannonball quickly disappeared when sculptors learned how to strengthen the internal structure of a statue with iron braces (iron being much stronger than bronze) . . .”

According to the paragraph above, sculptors in the Italian Renaissance stopped using cannonballs in bronze statues of horses because

- □ they began using a material that made the statues weigh less.
- □ they found a way to strengthen the statues internally.
- □ the aesthetic tastes of the public had changed over time.
- □ the cannonballs added too much weight to the statues.

At the end of the training, the participant was asked:

Do you have any questions or concerns regarding verbal report?

Are you ready to respond to each one of the two reading sets and report your response?

[The source of the reading mini-test is ETS (2003, p. 11)].

APPENDIX K

A Sample Transcript

RS#1	ID#20	AR#51	SR#24A		
Item					Transcript
#	Type	Timestamp	Codes	Item (+/−)	Notes: I did not look at the passage. I moved on to the questions directly. In my opinion, reading the passage on this kind of tests is timewasting.
1	BC-*v*	00:19	v1, v10, v8	−	I like these questions because they do not take much time, I mean, you do not have to read the whole paragraph or the passage to answer these questions. So here, I read the sentence and figured out what the word means in this sentence <v1>. You can figure out the meaning of the word by using the sentence. So here, I chose Option 3 "rapidly" because it sounds like 'he gained the power so fast' and "rapidly" means 'fast'; that is why I chose Option 3 "rapidly". I was afraid of Option 2 "greatly" because it has the same meaning, I mean, 'gaining the power so fast' as does "rapidly", but I did think that "greatly" would sound OK in this context; that is why I chose "rapidly" <v10>. I checked the other options, but they did not sound OK to me <v8>.
3	I-*rp*	03:06	rp2, rp5,	+	I think this was pretty much easy for me. The answer is Option 3 because the paragraph says that "bankers and investors manipulated the banking system for their own profit", which suggests that 'they do not care about the others'. ‘They want to become rich and not bring other people to share with them and … the economy system’ <rp2>. I eliminated Option 1 because it has the opposite meaning of what is stated in the paragraph. I also eliminated Option 4 because Andrew Jackson did not want to be favored by certain groups but not others <rp5>. He wanted to be fair and favored by all groups.

(Table continues.)

Appendix K: A Sample Transcript (continued)

7	BC-*fi*	09:00	fi1, fi6, fi4,	+	The given word "monopolies" helped me understand the paragraph a lot <fi1>. I read the options. I eliminated Option 1 because the paragraph says "Religion and politics, they believed, should be kept clearly separate …", because also the government is democratic so it aims to separate politics from religion. Option 3 "destroying monopolies" is one of the reasons but not 'the main reason'; it was part of the answer not the whole answer <fi6>. I chose Option 2 because it is more comprehensive <fi4>. It supports a person's freedom of choice of religion.

Notes. RS= reading set; ID=participant; AR=audio recording; SR=screen recording; Item (+/−)= item answered right or wrong. Strategy codes label highlighted segments in the transcripts.

APPENDIX L

Results of Item Analyses

	Item types	*P*	*Rpbi*	α (w/o)	Decision for item inclusion in data analysis
Reading Set 1	1. BC-*v*	0.68	0.35	0.867	Yes
	2. BC-*fi*	0.52	0.26	0.869	No
	3. I-*rp*	0.68	0.21	0.870	Yes
	4. BC-*fi*	0.72	0.40	0.865	No
	5. BC-*fi*	0.88	0.09	0.871	No
	6. BC-*v*	0.64	0.16	0.872	No
	7. BC-*fi*	0.64	0.60	0.860	Yes
	8. BC-*v*	0.84	0.35	0.867	No
	9. I-*bi*	0.24	0.08	0.873	Yes
	10. BC-*nf*	0.36	0.10	0.873	Yes
	11. BC-*ss*	0.48	0.21	0.871	Yes
	12. I-*it*	0.60	0.53	0.862	Yes
	13. R2L-*ps*	0.28	0.19	0.863	Yes
Reading Set 2	1. BC-*fi*	0.72	0.34	0.867	No
	2. BC-*v*	0.64	0.17	0.871	No
	3. I-*bi*	0.24	0.16	0.871	Yes
	4. BC-*v*	0.64	0.47	0.864	Yes
	5. BC-*fi*	0.76	0.52	0.863	Yes
	6. I-*rp*	0.76	0.59	0.861	Yes
	7. BC-*v*	0.60	0.38	0.866	No
	8. BC-*pr*	0.88	0.28	0.868	Yes
	9. BC-*ss*	0.60	0.49	0.863	Yes
	10. BC-*fi*	0.60	-0.05	0.877	No
	11. I-*it*	0.56	0.55	0.861	Yes
	12. R2L-*st*	0.30	0.32	0.859	Yes

Notes. p = item difficulty; r_{pbi} = item discriminability; α (w/o) = Cronbach's Alpha without the item.

APPENDIX M

Abbreviations of Task and Item Types

Used in Strategy Codes

Basic Comprehension (BC)

- BC-*v* (= vocabulary)
- BC-*fi* (= factual information)
- BC-*nf* (= negative fact)
- BC-*pr* (= pronoun reference)
- BC-*ss* (= sentence simplification)

Inferencing (I)

- I-*bi* (= basic inference)
- I-*it* (= insert text)
- I-*rp* (= rhetorical purpose)

Reading-to-Learn (R2L)

- R2L-*ps* (= prose summary)
- R2L-*st* (= schematic table)

VITA

Mohammed Assiri

Candidate for the Degree of

Doctor of Philosophy

Thesis: TEST-TAKING STRATEGY USE ON THE READING SECTION OF THE TOEFL IBT: A STUDY OF ARAB ESL LEARNERS

Major Field: English

Biographical:

Education:

Received Bachelor of Arts in English at King Khalid University, Abha, Saudi Arabia in 2000.

Received Master of Arts in Teaching English as a Second Language at the University of Kansas, Lawrence, Kansas in 2004.

Completed the requirements for Doctor of Philosophy in English at Oklahoma State University, Stillwater, Oklahoma in July, 2011.

Experience:

ESL Instructor at the English Language Center, Institute of Public Administration, Riyadh, Saudi Arabia (2004-2007)

Testing Coordinator at the English Language Center, Institute of Public Administration, Riyadh, Saudi Arabia (2004-2006)

ESL Teaching Assistant at the English Language Center, Institute of Public Administration, Riyadh, Saudi Arabia (2000-2002)

Professional Memberships:

International Language Testing Association (ILTA)

Teaching English to Speakers of Other Languages (TESOL)

Poetics and Linguistics Association (PALA)

Name: Mohammed Assiri Date of Degree: July, 2011

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: TEST-TAKING STRATEGY USE ON THE READING SECTION OF THE TOEFL IBT: A STUDY OF ARAB ESL LEARNERS

Pages in Study: 331 Candidate for the Degree of Doctor of Philosophy

Major Field: English

Scope and Method of Study: With the focus on how a sample of 25 Arab ESL learners respond to the TOEFL-iBT reading tasks, this study aimed to find out what strategies respondents tend to use, investigate if there are differences between high- and low-scorers in strategy use, and determine aspects of effective strategy use among respondents.

Data were collected using a procedural integration of stimulated recall, self-observation, and retrospective interview (SRSORI). A pilot study was conducted to evaluate and refine materials and procedures. Data collection was carried out over three stages. First, each participant was oriented to SRSORI and trained in producing verbal reports. Second, the participant responded to two reading sets, and thus was engaged in two SRSORI sessions. And third, the participant was debriefed about his test performance and scores. Data were then transcribed, and results of item analyses were used to decide on episodes for coding. A coding scheme was constructed for each task item by means of inductive coding. Data analysis made use of frequencies to identify patterns of strategy use and qualitative accounts to describe strategy use in relation to such factors as item format, scoring level, and answer correctness.

Findings and Conclusions: Major findings were as follows. Test takers used strategies depending on item format and difficulty in ways that allowed them to achieve different goals, adapt strategies to various task items, and apply strategies using several textual and technical means. High test performance and scoring were characterized by superior skills of both comprehension and test-management as well as high levels of strategic awareness and monitoring. Conversely, low test performance and scoring were associated with poor skills of comprehension and excessive use of test-wiseness. And, test takers sequenced strategies such that certain strategies derived from other strategies, endorsed or facilitated functions of other strategies, or acted in sync with other strategies.

It was concluded that strategy use has three facets: purposeful, multi-form, and resourceful. High test performance and scoring on reading tasks draw on response behaviors and strategies that differ from those associated with low performance and scoring. And, aspects of effective strategy use occur in tandem with logical sequences of strategies.

ADVISER'S APPROVAL: Dr. Gene Halleck