ISSN 1556-9012

ETS TOEFL

Monograph Series

MS - 33
April 2006

Strategies in Responding to the New TOEFL Reading Tasks

Andrew D. Cohen
Thomas A. Upton

Strategies in Responding to the New TOEFL Reading Tasks

Andrew D. Cohen
University of Minnesota

Thomas A. Upton
Indiana University—Indianapolis

RR-06-06

ETS is an Equal Opportunity/Affirmative Action Employer.

As part of its educational and social mission and in fulfilling the organization's non-profit Charter and Bylaws, ETS has and continues to learn from and also to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

To obtain more information about the TOEFL programs and services, use one of the following:

E-mail: toefl@ets.org
Web site: www.ets.org/toefl

Foreword

The TOEFL® Monograph Series features commissioned papers and reports for TOEFL 2000 and other Test of English as a Foreign Language™ (TOEFL) test development efforts. As part of the foundation for the development of the TOEFL Internet-based test (TOEFL iBT), papers and research reports were commissioned from experts within the fields of measurement, language teaching, and testing through the TOEFL 2000 project. The resulting critical reviews, expert opinions, and research results have helped to inform TOEFL program development efforts with respect to test construct, test user needs, and test delivery. Opinions expressed in these papers are those of the authors and do not necessarily reflect the views or intentions of the TOEFL program.

These monographs are also of general scholarly interest, and the TOEFL program is pleased to make them available to colleagues in the fields of language teaching and testing and international student admissions in higher education.

The TOEFL 2000 project was a broad effort under which language testing at ETS® would evolve into the 21st century. As a first step, the TOEFL program revised the Test of Spoken English™ (TSE®) and introduced a computer-based version of the TOEFL test. The revised TSE test, introduced in July 1995, is based on an underlying construct of communicative language ability and represents a process approach to test validation. The computer-based TOEFL test, introduced in 1998, took advantage of new forms of assessment and improved services made possible by computer-based testing, while also moving the program toward its longer-range goals, which included:

- the development of a conceptual framework that takes into account models of communicative competence
- a research program that informs and supports this emerging framework
- a better understanding of the kinds of information test users need and want from the TOEFL test
- a better understanding of the technological capabilities for delivery of TOEFL tests into the next century

Monographs 16 through 20 were the working papers that laid out the TOEFL 2000 conceptual frameworks with their accompanying research agendas. The initial framework document, Monograph 16, described the process by which the project was to move from identifying the test domain to building an empirically based interpretation of test scores. The subsequent framework documents, Monographs 17-20, extended the conceptual frameworks to the domains of reading, writing, listening, and speaking (both as independent and interdependent domains). These conceptual frameworks guided the research and prototyping studies described in subsequent monographs that resulted in the final test model. The culmination of the TOEFL 2000 project is the TOEFL iBT, which was introduced in September 2005.

TOEFL Program
ETS

Abstract

This study describes the reading and test-taking strategies that test takers used in the Reading section of the LanguEdge courseware (ETS, 2002a). These materials were developed to familiarize prospective respondents with the new TOEFL®. The investigation focused on strategies used to respond to more traditional single selection multiple-choice formats (i.e., Basic Comprehension and Inferencing questions) and the new selected-response (multiple selection, drag and drop) Reading to Learn items. The latter were designed to simulate the academic task of forming a comprehensive and coherent representation of an entire text, rather than focusing on discrete points in the text. Verbal report data were collected from 32 students, representing four language groups (Chinese, Japanese, Korean, and Other) doing the Reading section tasks from the LanguEdge courseware materials. Students were randomly assigned to two of the six reading subtests, each consisting of a 600–700 word text with 12–13 items. Subjects' verbal report accompanying items representing each of the ten item types was evaluated to determine strategy use. The findings provide insights into the response behaviors prompted by the reading tasks on the new TOEFL.

Key words: Academic reading comprehension, reading, reading strategies, test-taking strategies, verbal report, language use strategies

Acknowledgements

We express thanks to our research assistants—Mesut Akdere, Haesook Koo, Hitomi Maeda, and Hongzhuan Song—for their efforts in data collection, transcription, and coding, and to the 32 student participants. Special thanks are also given to Mesut Akdere, who in addition to his own RA duties, worked with and oversaw the efforts of the research assistants and greatly facilitated the execution of the study. Thanks also to the anonymous ETS reviewers, Mary Enright, and Mary Schedl for their constructive feedback on earlier drafts of this report.

Table of Contents

List of Tables

Page

Background

ESL Reading Comprehension

In the TOEFL® monograph *TOEFL 2000 Reading Framework: A Working Paper*, Enright et al. (2000) outline three main perspectives for understanding the nature of reading comprehension: the task perspective, the processing perspective, and the reader purpose perspective. In reviewing these three perspectives, Enright and Schedl (2000), in their ETS report *Reading for a Reason: Using Reader Purpose to Guide Test Design*, consider the reader purpose perspective, which "describes reading in terms of the coordinated application of knowledge and processes to a text or texts in the service of a goal or purpose," as representing the best model for assessment design (p. 4).

The reader purpose perspective recognizes that the reading process is very much an individual, cognitive process—what Bernhardt (1991) has called "an intrapersonal problem-solving task" (p. 6). From this perspective, task characteristics as well as reader's knowledge and personal abilities play a role in the degree of reading success. Performance variation in reading comprehension occurs due, to a large extent, to individual differences in linguistic knowledge and general and domain-specific background knowledge. Linguistic knowledge, often referred to as *bottom-up skills*, includes the ability to decode phonological, orthographic, lexical, syntactic, and discourse features of a text (Birch, 2002). There is little question that many poor readers are lacking in bottom-up abilities such as rapid, automatic word recognition. General and domain-specific background knowledge, often referred to as *top-down skills*, represents the content, social, and implicit knowledge that readers bring to a text (Bernhardt, 1991).

While linguistic and background knowledge appear to be primary sources for individual differences in reading skills, there are clearly numerous other variables (of unequal and variable importance) that can influence how first-language (L1) readers go about trying to understand an academic text and how successful those efforts will be. Enright et al. (2000) note that these variables include: cognitive processing abilities (e.g., working memory efficiencies), text type (e.g., expository vs. narrative), reading task, strategy use, affect (e.g., motivation, anxiety), topic, and L1, among others. It is the interplay between all these variables that influences how individual respondents perform on given reading tasks as they seek to achieve a particular goal or purpose.

ESL Reading and Test-Taking Strategies

The identification of strategies used by test takers on tests of reading comprehension is of interest. There has been a growing recognition of the importance of gaining a better understanding of how reading and test-taking strategies are used on tests as part of the process of construct validation: "the relationship between test performance and the construct, or ability, it is intended to measure" (Anderson, Bachman, Perkins, & Cohen, 1991, p. 42). In short, as Cohen (1994a) has noted, "In order to assess reading comprehension in a second or foreign language, it is necessary to have a working knowledge of what that process entails" (p. 211). As Bachman and Palmer (1996) more recently have stated, "Unless we can demonstrate that the inferences [about language ability] we make on the basis of language tests are valid, we have no justification for using test scores for making decisions about individuals . . . we must demonstrate that these inferences are appropriate for the decisions we need to make" (p. 95). Consequently, it is important to have good insight into what it is people who take reading comprehension tests do in order to complete them.

Reading Strategies

According to Carrell and Grabe (2002), it is clear that when reading,

> a reader engages in processing at the phonological, morphological, syntactic, semantic and discourse levels, as well as engages in goal setting, text-summary building, interpretive elaborating from knowledge resources, monitoring and assessment of goal achievement, making various adjustments to enhance comprehension, and making repairs to comprehension processing as needed. (p. 234)

While much of the reading process is automatic in nature—which is defined as reading skill (Williams & Moran, 1989, p. 223)—and beyond our conscious control, readers do exert a significant level of active control over their reading process through the use of strategies, which are conscious procedures that are deliberate and purposeful (Williams & Moran, 1989, p. 98; Urquhart & Weir, 1998). While *processes* are general, subconscious or unconscious, and more automatic, *strategies* are subject to control, more intentional, and used to act upon the processes (see Cohen, 2005).

In keeping with the understanding of reading as a problem-solving process, reading strategy analysis provides insights as to how readers interact with the text and how their choice

of strategies influences their comprehension of the text. A focus on reading strategies helps researchers determine the extent to which readers genuinely understand the purpose of what they are reading, how they go about making sense of what they read, and what they do when they do not understand some aspect of the text.

Strategy use and efficacy are clearly influenced by the proficiency level of the reader. For example, Carrell and Grabe (2002) have underscored the importance of vocabulary knowledge in dealing with second-language (L2) text, as well as noting the difficulty nonnative readers may have guessing words from context (a reading strategy), especially when the context is not very helpful. Furthermore, Skehan (1998) and others have warned that the task itself may pose true hurdles for the nonnative reader, and in fact, the length of a text and the nature of the questions asked about it on a test can may have a real impact on the strategies that L2 readers use and the results.

In their book, *Verbal Protocols of Reading: The Nature of Constructively Responsive Reading*, Pressley and Afflerbach (1995) set out to "identify and describe exhaustively the many processes, including strategies and responses, that readers carry out consciously as they read" (p. 15). Scouring the literature for studies using the think-aloud protocols that best reflect the conscious, online reading processes of readers, they reviewed 38 widely diverse primary research studies to develop and categorize a complete and coherent list of more than 150 conscious activities (i.e., reading strategies) that readers may use when working to comprehend a text. They grouped these reading strategies into three broad categories: (a) planning and identifying strategies that help in constructing the meaning of the text, (b) monitoring strategies that serve to regulate comprehension and learning, and (c) evaluating strategies by which readers reflect or respond in some way to the text.

Research in second-language reading has shown that second-language readers draw on this same array of reading strategies (e.g., Upton & Lee-Thompson, 2001; Carrell & Grabe, 2002). Planning and identifying strategies include planning how to read the passage, checking out any notable discourse features, remembering why the text is being read, checking for prior knowledge of the topic, reading through the whole text, reading portions selectively, looking for markers of meaning, and identifying portions of the text that are unclear. The strategies used during the reading process include monitoring ongoing understanding of the text, predicting what will come, rereading for clarification, and modifying comprehension based on new information.

In addition, there are strategies associated with identifying the discourse structure of the text and important information found in it: identifying key words and sentences carrying the main ideas and determining how the different parts of the text function in the discourse. Likewise, readers need to strategize when the reading remains unclear or ambiguous by inferring the meaning of words and ideas based on clues to meaning in the text. While this depiction of reading strategies is still only partial, it encompasses many of the more prominent strategies in the literature.

Test-Taking Strategies

Test-taking strategies are defined as those test-taking processes that the respondents have selected and are conscious of, at least to some degree. In other words, the notion of strategy implies an element of selection. Otherwise, the processes would not be considered as strategies. At times, these strategies constitute opting out of the language task at hand (e.g., through a surface matching of identical information in the passage and in one of the response choices). At other times, the strategies may constitute shortcuts to arriving at answers (e.g., not reading the text as instructed but simply looking immediately for the answers to the given reading comprehension questions). In such cases, the respondents may be using *test-wiseness*[1] to circumvent the need to tap their actual language knowledge or lack of it. This is consistent with Franson's (1984) assertion that respondents may not proceed via the text but rather around it. In the majority of testing situations, however, test-taking strategies do not lead to opting out or to the use of short cuts. In some cases, quite the contrary holds true. One second-language respondent (Hebrew) in a study of test-taking strategies in Israel determined that he had to produce a written translation of a text before he could respond to questions dealing with that text (Cohen & Aphek, 1979).

At times, the use of a limited number of strategies in a response to an item may indicate genuine control over the item, assuming that these strategies are well-chosen and are used effectively. At other times, true control requires the use of a host of strategies. It is also best not to assume that any test-taking strategy is a good or a poor choice for a given task. It depends on how given test takers—with their particular cognitive style profile and degree of cognitive flexibility, their language knowledge, and their repertoire of test-taking strategies—employ these strategies at a given moment on a given task. Some respondents may get by with the use of a limited number of strategies that they use well for the most part. Others may be aware of an extensive number of strategies but may use few, if any of them, effectively. So, for example,

while a particular skimming strategy (such as paying attention to subheadings) may provide adequate preparation for a given test taker on a recall task, the same strategy may not work well for another respondent. It also may not work well for the same respondent on another text that lacks reader-friendly subheadings.

The ability of learners to use language strategies has been referred to as their *strategic competence*—a component of communicative language use (Canale & Swain, 1980). This model puts the emphasis on *compensatory* strategies—that is, strategies used to compensate or remediate for a lack in some language area. Bachman (1990) provided a broader theoretical model for viewing strategic competence, and then Bachman and Palmer (1996) refined the Bachman (1990) categories somewhat. Their current framework includes: an assessment component, whereby the respondents (in this case of language testing) assess which communicative goals are achievable and what linguistic resources are needed; a goal-setting component, wherein the respondents identify the specific tasks to be performed; a planning component, whereby the respondents retrieve the relevant items from their language knowledge and plan their use; and an execution component, whereby the respondents implement the plan. Hence, this latest framework for strategic competence is broad and includes test-taking strategies within it. It would be interesting to see the extent to which respondents on the new TOEFL are doing initial assessment, then goal setting and planning, as opposed to simply executing or responding to the task.

As long as the task is part of a test, students may find themselves using strategies that they would not use under pretest conditions. It is for this reason that during the pilot phase, it is crucial for test constructors to find out what their tests are actually measuring.

Testing Academic Reading Comprehension

Enright & Schedl (2000) note that "in large-scale assessment, the goal is to gather evidence of individual differences in critical competencies that contribute to different levels of performance in a domain" (p. 5). For the new TOEFL (originally known as TOEFL 2000 and next generation TOEFL), as noted above, reader purpose was selected as the basis for reading comprehension assessment. Since the new TOEFL is intended to "measure examinee's English-language proficiency in situations and tasks reflective of university life in North America" (Jamieson, Jones, Kirsch, Mosenthal, & Taylor, 1999, p. 10), the reading section of this test is designed to simulate the types of reading tasks that students are expected to do in university-

level academic settings. Consequently, it was originally proposed that the new TOEFL should evaluate academic reading comprehension and focus on four main purposes "identified as likely to be important in academic contexts" (Enright & Schedl, 2000, p. 5):

1. Reading to find information
2. Reading for basic comprehension
3. Reading to learn
4. Reading to integrate information across multiple texts

The most recently available new TOEFL reading task specifications (ETS, 2003), which were derived from the work of Enright et al. (2000) and Enright and Schedl (2000), evolved somewhat with the resulting reading tasks designed specifically to focus on "reading for basic comprehension tasks, reading to learn tasks, and a third group of tasks, inferencing tasks, that have elements of both of the other types" (p. 2). According to the task specifications (ETS, 2003):

> Basic comprehension questions are used to assess lexical, syntactic, and semantic abilities and the ability to understand important information presented in sentence-level propositions. . . . Reading to learn is seen as requiring additional abilities beyond those required for basic comprehension. Reading to learn questions assess specific abilities that contribute to learning including the ability to recognize the organization and purpose of a text, to distinguish major from minor ideas and essential from nonessential information, to conceptualize and organize text information into a mental framework, and to understand rhetorical functions such as cause-effect relationships, compare-contrast relationships, arguments, and so on. . . .
>
> Inferencing tasks . . . shares [sic] some characteristics with basic comprehension tasks and some characteristics of reading to learn tasks. Inference questions, insert text questions, and rhetorical purpose questions were originally considered basic comprehension tasks because they were item types associated with the old TOEFL test, where the focus of the items was primarily on lexical, syntactic, and semantic abilities and they were usually based on single sentences or small portions of the text. However, after redefining these tasks for the Next Generation TOEFL . . . they are now characterized by some important reading to learn features. While they can still be used to

> test sentence-level information, as basic comprehension items do, they can also be used to test information across larger areas of the text. They may also require the abilities associated with reading to learn, including connecting information and recognizing the organization and purpose of the text. (pp. 2–3)

It is these three broad categories of reading tasks (basic comprehension, reading to learn, and inferencing)—each with multiple types—that form the basis for the reading portion of the new TOEFL.[2, 3] The new TOEFL draws on a total of ten item types (described in more detail in the Methods section) representing five Basic Comprehension tasks, three Inferencing tasks, and two Reading to Learn tasks.

In addition, text length and text type are important considerations in the assessment of reading comprehension, especially when the goal is to specifically assess academic reading ability. The Reading section of the new TOEFL incorporates fewer but longer (600–700 vs. 300–400 words) texts than used in previous TOEFL test designs (i.e., the traditional paper-based TOEFL and the newer, computer-based test, TOEFL CBT). The reasons given for this are that longer texts better represent the academic experiences of students and that they better facilitate the development of reading to learn purposes in the test design (Mary Schedl, personal communication, April 6, 2004). With regard to text type, previous TOEFL reading passages "consisted primarily of a particular type of expository text in which a number of discrete facts are loosely integrated and developed" (ETS, 2003, p. 1). Along with expanded length, the texts in the Reading section of the new TOEFL (each test has three texts on different general academic topics) include a broader selection of academic text types, classified by author purpose: (a) exposition, (b) argumentation, and (c) historical biographical/autobiographical narrative. Each of these has at least one structure, such as classification, comparison/contrast, cause/effect, and problem/solution, with information presented from more than one perspective or point of view (ETS, 2003).

In sum, it is the belief that manipulation of reading task, along with text length and text type, allows for the fair, reliable, and efficient evaluation of individual academic reading proficiency while reflecting "the conditions under which college and graduate students read in formal academic settings" (Enright et al., 2000, p. 32).

Design of Reading Test Items

Computerized, large-scale assessments of reading ability offer both opportunities and constraints in the item types used. Primary constraints revolve around issues of fairness. In order to be "fair to examinees from a wide variety of cultural and educational backgrounds and to be reasonably priced so that access to education does not depend on income," test items have to be designed so that they are not only "suited to the examinee population," but are also economical in terms of time and cost (Enright & Schedl, 2000, p. 30). For these reasons, paper-based tests have relied almost exclusively on single-selection multiple-choice questions; computerized reading tests for the most part continue to rely on multiple-choice question item types for the same reasons. Computerized tests have other potential shortcomings, not least of which is the fact that "the experience of reading online is generally less appealing than that of reading paper copy" (Enright et al., 2000, p. 41). Lack of experience with essential computer skills (e.g., using a mouse, scrolling text, clicking on text, keyboarding) can also be issues in designing fair computerized tests.

Nevertheless, computerized tests offer advantages and opportunities for "novel types of selected responses," providing a wider variety of response formats than can be used on paper-and-pencil exams (Mary Schedl, personal communication, April 6, 2004). For example, in response to a prompt, test takers can be directed to point and click on relevant portions of the text, or point-click and drag chunks of text from one part of the screen to another (Enright et al., 2000, p. 32). While these more novel response options can still be used within the single-selection multiple-choice format, computerized tests also permit the development of questions with what are referred to as multiple-selection multiple-choice responses. The new TOEFL designers have selected this type of item structure as appropriate for assessing Reading to Learn ability.

LanguEdge courseware, introduced in 2002 to acquaint test users with the new TOEFL tasks, included two prototype test forms. The Reading sections of these prototype test forms included examples of these novel multiple-selection multiple-choice reading items—prose summaries and schematic tables.[4] For the prose summary, test takers are asked to "complete a summary of a text, one or two sentences of which are provided" by selecting three additional sentences from a list of six that express the most important ideas in the passage (Enright & Schedl, 2000, p. 19). Distracters include ideas that either are not presented in the passage or are deemed as minor ideas. For the schematic table, test takers must "click and move sentences or

phrases into a table to complete a schematic representation [of the passage]. A correctly completed table should reveal an integrated mental model of how the two dimensions fit together conceptually based on the information in the text" (Enright & Schedl, 2000, p. 19). The focus of both of these multiple-selection multiple-choice items is on the ability to identify major ideas and important information in a text. The value of these questions is greater than the typical single-response questions, and partial credit is awarded if only part of the response is correct.

Last, it should be noted that the difficulty level of reading-comprehension tasks can be manipulated using one or more of (at least) three variables: type of information, type of match, and plausibility of distracters (Kirsh & Mosenthal, 1990; Mosenthal & Kirsch, 1992). These variables are carefully considered in the development of items for the new TOEFL.

In order to understand the demands placed on readers taking a reading comprehension test, one must understand the nature and demands of the question items themselves. It is in this context that an interpretation of the reading comprehension and test-taking strategies employed by individuals must be interpreted.

Purpose of This Study

Since the 1980s, there has been a call for the development of language tests that provide a better fit between "the tester's *presumptions* about what is being tested and the *actual* processes that the test taker goes through" (Cohen, 1984, p. 70). The goal is to have tests that fairly reflect the language competence of the test taker to handle the language task being evaluated, while guarding against opportunities for selecting the right answers for the wrong reasons. It is no secret that test takers have developed numerous techniques for finding correct answers to reading tests "without fully or even partially understanding the text" (Cohen, 1986, p. 132). In other words, in a test claiming to evaluate academic reading ability, the premium needs to be on designing tasks calling for test takers to actually use academic reading skills in responding to items, rather than being able to rely on test-wiseness tricks. The new TOEFL makes great efforts to evaluate the reading comprehension abilities of test takers dealing with academic-like texts. It is clear that these reading tasks are able to discriminate better readers from poorer readers. How exactly do test takers go about trying to complete the tasks? What reading and test-taking strategies do students use to complete the tasks? Do different item types require different types of reading and/or test-taking strategies? At least one previous study (Anderson et al., 1991) has

suggested that item type can have a considerable influence on the reading and test-taking strategies that are used and on how they are used as well.

The purpose of this study is to describe the reading and test-taking strategies that test takers use to complete the reading tasks in the Reading sections of the LanguEdge courseware (ETS, 2002a) materials developed to introduce the design of the new TOEFL. In particular, it seeks to determine if there is variation in the types of strategies used when answering the three broad categories of question types, including the more traditional single-selection multiple-choice formats, which are used for Basic Comprehension and the Inferencing questions,[5] as well as the new selected-response (multiple-selection multiple-choice) Reading to Learn. As already stated, the Reading to Learn items were designed to assess an examinee's ability to form a more comprehensive and coherent representation of the whole text rather than focusing on discrete points in the text. In the LanguEdge prototype test, these three categories of items are further broken down into subtypes, with Basic Comprehension items including vocabulary, pronoun reference, sentence simplification, factual information, and negative fact items. Inferencing items deal with inference in general, rhetorical purpose, and the insertion of text. As already noted, Reading to Learn items include both what is referred to as prose summary and also schematic table items (see Testing Materials). It is expected that the findings will help provide a better understanding of the demands that the reading tasks for the new TOEFL place on students and the strategies that they use to respond to these demands.

Research Questions

As noted above, this study seeks to determine if there is variation in the types of strategies used when answering the three broad categories of question types, including the more traditional single-selection multiple-choice formats, which are used for Basic Comprehension and Inferencing questions, as well as the new selected-response (multiple-selection multiple-choice) Reading to Learn items. The following is the guiding research question: What processing strategies do respondents use in producing answers to the Basic Comprehension, Inferencing, and Reading to Learn items on the reading subtest of the new TOEFL? More explicitly:

1. What reading strategies and test-taking strategies do respondents report using? Specifically, what strategies are used to complete each of the ten different test item types?

2. Do the Inferencing and the Reading to Learn items require and assess different academic-like approaches to reading than the Basic Comprehension questions, and are they more difficult?

The ETS TOEFL committee refers to academic texts as being—among other things—longer (600–700 vs. 300–400 words), fairly decontextualized, focused on broad academic topics, and organized using certain rhetorical features (Enright et al., 2000). Academic reading, then, is reading that is undertaken for the purpose of processing and understanding these academic texts, which in turn would require strategies for retaining ideas in working memory. Likewise, the Reading to Learn and Inference items were expected to call for the academic skills of identifying logical connectors and other markers of cohesion and determining how sections of passages interrelate in an effort to establish passage coherence. The claim is not that these skills are *exclusive* to academic reading, only that these are skills that effective academic readers are able to mobilize through their use of strategies. The question, in other words, is whether the Inferencing and Reading to Learn items do, in fact, prompt the use of such discourse-level strategies more than do the Basic Comprehension items, and whether these items are commensurately more difficult to answer.

Methodology

Verbal Reports

The act of reading a text to understand it and that of taking a test about the text involve cognitive processes that are not for the most part observable, nor are the strategies deployed by the reader/test taker for the purpose of accomplishing these activities. Consequently, in order to get the best picture possible of what readers do as they read test prompts and respond to test questions, verbal protocols are typically an instrument of choice. Indeed, verbal report as a means to investigate cognitive processes is fairly well established in many fields, most notably in psychology and education. In second-language acquisition studies, verbal report has been used to investigate the cognitive strategies of adult learners and children reading L2 texts (e.g., Hosenfeld, 1984; Block, 1986; Cavalcanti, 1987; Kern, 1994), writing in the L2 (e.g., Zamel, 1983; Raimes, 1987; Cohen & Cavalcanti, 1987, 1990; Skibniewski, 1990), and taking tests (e.g., Nevo, 1989; Anderson, 1991; Stemmer, 1991; Brown, 1993; Warren, 1996), among other things.

Green (1998) provides a comprehensive and in-depth overview of how verbal reports can be used in language testing. According to Green, "verbal protocols are increasingly playing a vital role in the validation of assessment instruments and methods [in that they] offer a means for more directly gathering evidence that supports judgments regarding validity than some of the other more quantitative methods" (p. 3). Green, in fact, notes that verbal reports are frequently used to address what he terms "one of the most fundamental questions" about language tests (p. 3): What is it that a test actually measures?

As noted by Leow and Morgan-Short (2004), however, "all verbal reports are not equal" (p. 36). Verbal reports can be elicited in different ways. One important variable is whether the verbal report is introspective or retrospective in nature. Introspective verbal reports are those gathered while subjects are performing a given task; retrospective verbal reports are those gathered after (usually immediately after) the completion of a given task. Introspective reports are considered more desirable since they are not limited by memory constraints or open to reconstructive modifications. Yet a caveat regarding introspective reports is that the process of collecting them may alter, to some extent, the task being undertaken because "the think-aloud data collection method itself acts as an additional task that must be considered carefully when examining learner performance" (Jourdenais, 2001, p. 373). Furthermore, verbalizations can be elicited that are either mentalistic—where subjects report on what they think they are doing to accomplish a task, or nonmentalistic—where subjects focus on the task and merely voice their thoughts without trying to explain what they are doing. Cohen (2000) has described this distinction as self-observational versus self-revelational, with the latter being seen as more accurately reflecting the actual thought processes.

As noted above, there is much evidence that verbal reports, especially those that are introspective and nonmentalistic in nature, provide significant insight into the cognitive processes of second-language learners. There is also strong evidence that the way subjects approach tasks while thinking aloud is very much in keeping with how they approach tasks normally (see Ericsson & Simon, 1993). To cite one recent study, Leow and Morgan-Short (2004) found that introspective, nonmentalistic verbal reports did not have a detrimental effect on the second language reading comprehension of the adult readers they studied.

Cohen (1986, 1991) divides verbal reports into three broad types. First is *self-report*, where learners provide descriptions of what they do, characterized by generalized statements, for

example, about their test-taking strategies. For example: "On multiple-choice items, I tend to scan the reading passage for possible surface matches between information in the text and that same information appearing in one of the alternative choices." Self-reports are retrospective and mentalistic in nature. Second is *self-observation*, which is the inspection of specific, not generalized, language behavior, either introspectively (i.e., within 20 seconds of the mental event) or retrospectively. For instance: "What 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." Self-observation is mentalistic in nature, but can be either introspective or retrospective. The third verbal report is *self-revelation*, what is typically called *think-aloud*. This is the stream-of-consciousness disclosure of thought processes while the information is being attended to. For example: "Hmm . . . I wonder if the information in one of these alternative choices also appears in the text." Self-revelation is introspective and nonmentalistic in nature, and is seen as most accurately reflecting learners' cognitive processes (Cohen, 2000; Ericsson & Simon, 1993). Verbal reports can and usually are comprised of some combination of these (Radford, 1974; Cohen & Hosenfeld, 1981; Cohen, 1987), so care must be taken in carefully controlling for the type of data collected.

Questionnaires and other kinds of prompts that ask learners to describe the way they usually take a certain type of language test are likely to elicit self-report data. Self-observation data would entail reference to some actual instance(s) of language testing behavior. For example, recollections of why certain distracters were rejected in search of the correct multiple-choice response on previously answered items would count as retrospective self-observation. Self-revelation or think-aloud data are only available at the time that the language event is taking place (i.e., within 20 seconds of it), and the assumption would be that the respondent is simply describing, say, the struggle to determine which five out of seven or more statements constitute the best set of main points for a text. Any thoughts that the respondent has that are immediately analyzed would constitute introspective self-observation. For example: "Now, does this utterance call for the present or imperfect subjunctive? Let me see . . . "

By asking test takers to think aloud as they work through a series of test items, it becomes possible to analyze the resulting protocol in order to identify the cognitive processes involved in carrying out the task. Think-aloud protocols have the advantage of giving a more direct view of how readers process a text as they indicate what they are doing at the moment they are doing it

(Cohen, 1987). Retrospective interviews, in turn, provide an opportunity for investigators to ask directed questions to gain clarification of what was reported during the think aloud.

Early work in verbal report with language testing found, for example, that some assumptions were ill founded. One was that technical vocabulary does not cause as much difficulty as nontechnical vocabulary and nontechnical vocabulary used technically within a given field. Furthermore, seemingly obvious discourse markers may not be so obvious to the L2 reader. In addition, the problems arising from syntactic features may be quite limited in scope—stemming mostly from structures such as heavy noun phrases (Cohen, Glasman, Rosenbaum-Cohen, Ferrara, & Fine, 1979). Cohen (1986) laid out a series of measures to be taken to ensure that verbal report tasks could be used effectively to obtain data on the reading process.

More recently, numerous studies have been conducted to determine the strategies that students use to read texts (see Singhal, 2001, for a review). Upton (1997, 1998), for example, reported on 11 native speakers of Japanese—one half still taking ESL classes and one half having finished with courses. They were asked to provide think-aloud protocols while they read academic passages. In retrospective interviews, they then listened to their tape-recorded protocols and were asked to clarify and explain their thoughts. Upton's study demonstrated how verbal report can be used to describe the ways in which nonnative speakers of a language can misconstrue the meaning of words and phrases as they read an L2 text, and how this throws off their understanding of the entire text. He found that many reading errors could be explained in terms of what Laufer (1991) has called *synforms*—that is, words that look or sound similar to other words that the readers know. The respondents would make the vocabulary in the passage conform in their minds to what they thought the meaning of these look-alike words was. A more recent study by Upton and Lee-Thompson (2001) used verbal report with 20 native speakers of Chinese and Japanese to explore the question of when and how they use L1 resources while reading L2 texts.

Verbal report measures have helped determine how respondents actually take reading comprehension tests as opposed to what they may be expected to be doing (Cohen, 1984; 1994a, pp. 130–136). Studies calling on respondents to provide immediate or delayed retrospection as to their test-taking strategies regarding reading passages with multiple-choice items have, for example, yielded the following results:

- Whereas the instructions ask students to read the passage before answering the questions, students have reported either reading the questions first or reading just part of the article and then looking for the corresponding questions.
- Whereas advised to read all alternatives before choosing one, students stop reading the alternatives as soon as they have found one that they decide is correct.
- Students use a strategy of matching material from the passage with material in the item stem and in the alternatives, and prefer this surface-structure reading of the test items to one that calls for more in-depth reading and inferencing.
- Students rely on their prior knowledge of the topic and on their general vocabulary.

From these findings and from others, there is a description emerging of what respondents do to answer questions. Unless trained to do otherwise, they may use the most expedient means of responding available to them, such as relying more on their previous experience with seemingly similar formats than on a close reading of the description of the task at hand. Thus, when given a passage to read and multiple-choice items to answer, they may attempt to answer the items just as they have answered other multiple-choice reading items in the past, rather than paying close attention to what is called for in the current one. Often, this strategy works, but on occasion the particular task may require subtle or major shifts in response behavior in order to perform well.

With regard to the training of respondents in how to provide verbal report, early descriptions of verbal report methods usually included the stipulation that respondents not be given instructions as to what to provide verbal reports about. They were to be left to their own devices since any instructions might lead to biased processing. But anyone who has been faced with analyzing transcriptions of undirected verbal report protocols has seen how such data are likely to be too general and incomplete. So, even methodological hard-liners like Ericsson and Simon are in favor of instructions to the respondents so as to make the verbal reports more complete (1993, p. 11). Green (1998; see also Cohen, 2000) underscores the need to ensure that "as much valid and complete data as possible are collected," and so highlights the importance of training subjects to give quality verbal reports (p. 41). In his summary of key steps in collecting verbal report data, Green includes the following:

- Prepare the subjects: Give "clear and unambiguous instructions" (p. 41)

- Brief the subjects: “Brief individuals on what is required of them and explain the procedure that is to be used” (p. 41)
- Practice the technique: “Give individuals some practice tasks to familiarize them with the technique” (p. 42)
- Practice the task: “Provide individuals with practice in the chosen procedure on the task . . . that will form the focus for the study” (p. 42)
- Give feedback: “Provide feedback on thinking aloud . . . performance” (p. 42)

Not so surprisingly, then, many studies now do include instructions to elicit particular cognitive behaviors. For example, reading researchers have cued different processes in the different studies. Pressley and Afflerbach (1995) found one study in which the respondents were informed about the importance of summarization and then were directed to create a summary of what they read. A second study asked respondents to attend to content and style when reading, while other studies required subjects to make inferences. The authors concluded that prompting respondents to use particular processes may be necessary: “. . . [I]t is reasonable to prompt [processes] in order to assure that a sample of the target processes will, in fact, be observed” (p. 133). Not only has it proven effective to have respondents receive specific prompts as to what to verbal report about, but it has also been seen that instruction in how to provide verbal report for a given task improves the quality of the data. Ericsson and Simon (1993) found that, to assure that the verbal report does not interfere with the task at hand, there must be warm-up trials (after the instructions) with tasks that yield easy-to-analyze think-aloud, introspective, and retrospective reports.

Pilot Study

Because of the qualitative nature of the study, a pilot study was run in order to test and refine the various instruments, procedures, and rubrics to be used, and to carefully train and “calibrate” the research assistants (RAs) in the verbal report process and the coding of the data. The pilot study consisted of four subjects, representing four different language backgrounds. Video tapes of the training sessions were used for review, evaluation, and calibration of the research assistants to help ensure reliability across language groups. Coding rubrics and other procedures were also refined and are described in this report. Modifications to the final study based on the pilot study are described in the following respective sections.

Data Collection for the Main Study

Recruitment of Test Takers

Based on the pilot study, students still enrolled in preacademic ESL courses at the University of Minnesota found the reading and test-taking demands of the Reading sections of the LanguEdge courseware (ETS, 2002a) materials well beyond their language abilities. As a result, a large portion of the strategies these students used were either sheer guesses or based on perceived test design clues and assumptions (i.e., test-wiseness strategies), primarily because they were unable to understand large sections of the texts, or even overall main ideas. Since the number of students included in the study was limited, and the purpose was to evaluate the reading and test-taking strategies used by students as they worked to answer the reading items (and the evaluation of strategies that involved primarily guessing and prereading-based assumptions were not considered informative), it was decided that only students who had a minimum scaled score of 11 (out of 25 possible) for the Reading Test section of the timed version of the LanguEdge test, which was used as the proficiency indicator, would be included in the main study. A scaled score of 11 represents the 34th percentile; subjects scoring below this fall into the bottom one third of examinees making up the Field Study sample (ETS, 2002b). By restricting study participants only to those who had a minimum score of at least 11, we were able to ensure that our subjects would be at a proficiency level where they would have at least a basic level of understanding of the test passages and so would provide a more accurate representation of what test takers do to answer reading test items using strategies other than guessing. Students who scored below 11 generally found the reading and test tasks beyond their abilities and resorted primarily to guessing and test-wiseness strategies in attempting to answer the questions.

Thirty-two nonnative speakers of English representing four language groups (Chinese, Japanese, Korean, and Other) were recruited to participate in the study. Flyers of introduction and invitation were sent to groups of suitable students attending the University of Minnesota, in Minneapolis, MN (see Appendix A). Table 1 provides a description of the students, including their study ID number, first language (L1), current education status, discipline of study, and length of residence (LOR) in the United States—as determined by the background questionnaire filled out by each participant at the start of the study (see Appendix B)—as well as their reading score on the reading section of the pretest version of the LanguEdge courseware materials.

The mean score for the 32 subjects on the timed LanguEdge reading pretest was 18.9 (out of 25), which places them at about the 75th percentile in relation to ETS' 2002 Field Study (ETS, 2002b). The mean scores for the subjects by language group were: Chinese subjects, 18.1 (~70th percentile); Japanese subjects, 18.6 (~75th percentile); Korean subjects, 17.4 (~66th percentile); and Other subjects, 21.5 (~90th percentile).

Table 1

Test-Taker Characteristics

ID #	L1	Sex	Age	Education status	Discipline	LOR[a]	Pretest score[b]
Chinese							
C1	Chinese	M	30	Graduate	Biostatistics	36 mths	17
C2	Chinese	F	25	Graduate	Statistics	24 mths	21
C3	Chinese	F	37	Graduate	Educational policy/admin	6 mths	21
C4	Chinese	M	28	Graduate	Computer science	60 mths	15
C5	Chinese	M	24	Graduate	Mathematics	20 mths	17
C6	Chinese	M	29	Graduate	Electrical engineering	48 mths	18
C7	Chinese	F	25	Graduate	Statistics	36 mths	17
C8	Chinese	F	24	Undergraduate	Economics	10 mths	19
Japanese							
J1	Japanese	F	21	Undergraduate	Linguistics	5 mths	20
J2	Japanese	F	22	Undergraduate	English	6 mths	15
J3	Japanese	F	23	Undergraduate	Civil engineering	4 mths	15
J4	Japanese	M	20	Undergraduate	Mechanical engineering	4 mths	22
J5	Japanese	M	34	Graduate	Business	19 mths	20

(Table continues)

Table 1 (continued)

ID #	L1	Sex	Age	Education status	Discipline	LOR[a]	Pretest score[b]
J6	Japanese	F	22	Graduate	Educational psychology	8 mths	21
J7	Japanese	F	33	Graduate	Public policy	29 mths	20
J8	Japanese	F	24	Graduate	(Not given)	19 mths	16
Korean							
K1	Korean	M	28	Graduate	Economics	5 mths	17
K2	Korean	F	22	Undergrad	Global studies	48 mths	20
K3	Korean	F	20	Undergrad	Biochemistry	48 mths	20
K4	Korean	M	28	Undergrad	Economics	60 mths	11
K5	Korean	M	27	ESL-only	ESL	9 mths	15
K6	Korean	F	30	Graduate	Education	48 mths	20
K7	Korean	M	20	Undergrad	Marketing	36 mths	18
K8	Korean	F	25	Graduate	Computer science	18 mths	18
Other							
O1	Turkish	M	26	Graduate	Civil engineering	48 mths	22
O2	Turkish	M	26	Graduate	Mathematics	48 mths	23
O3	Thai	F	23	Graduate	Applied linguistics/TESL	24 mths	20
O4	Bengali	M	27	Graduate	Finance	84 mths	23
O5	Arabic	M	22	Undergrad	Mechanical engineering	40 mths	20
O6	Turkish	M	25	Graduate	Mathematics	36 mths	22
O7	Arabic	M	24	Undergraduate	Undeclared	36 mths	21
O8	Turkish	M	26	Graduate	Curriculum and instruction	24 mths	21

[a]LOR = length of residence in the U.S. [b]Timed reading section from the LanguEdge prototype of the new TOEFL. Perfect scaled score = 25; scores were scaled based on the test form used by each subject (see Table 2).

To encourage participation in this fairly long and cognitively demanding study, students were offered a stipend equivalent to $15/hour for their time. At the time of recruitment, students were given information about the aims of the study, and in keeping with IRB requirements at the University of Minnesota, asked to sign a consent form (see Appendix C).

Test Materials

The LanguEdge courseware Reading section from one of the two published forms on the CD-ROM served as the pretest before the protocol portion of the study was conducted. Because LanguEdge tests were administered to a generally representative sample of the TOEFL population, the use of the LanguEdge reading test as a pretest ensured that performance could be related to levels of proficiency in the TOEFL population. Since there are two forms of the LanguEdge tests, respondents received in a counterbalanced fashion one as the pretest and the other as the form for which a protocol was obtained. The form that would serve as the pretest and the one that would provide the study prompts were assigned to each language group prior to subject recruitment (see Table 2).

A further rationale for administering a form of the LanguEdge reading section as the pretest was that it gave the test takers exposure to and practice in doing the task types on the reading tasks from the LanguEdge test before they were asked about the strategies they had used. A brief orientation to the general test types with examples was also provided through the LanguEdge test material before respondents began the placement test. Both the orientation and the pretest familiarized the study participants with the nature of the LanguEdge reading section and its format prior to the data collection stage, and gave them practice in responding to the test prior to the collection of data on their performance. This procedure is consistent with the reality that in a high-stakes environment, test takers familiarize themselves with what will be on the test.

Reading Test

The LanguEdge courseware (ETS, 2002a) materials include two prototype forms of the Reading section of the new TOEFL. As previously noted, subjects were assigned one or the other form, under regular time constraints, as a pretest to determine general reading proficiency and to familiarize them with the test design and expectations. Each pretest included three sets, each one consisting of a ~600–word text with 12–13 test items accompanying it. Subjects had 25 minutes to complete each of the three sets. For the data collection stage (outlined below), subjects

completed two of the three sets of texts and test items under no time constraint in order to facilitate the verbal report process. Each set included test items designed to evaluate reading for Basic Comprehension, Inferencing, and Reading to Learn abilities, which are detailed in the following section. Table 2 outlines the test forms taken by subjects during the pretest and study stages, in which two of the three passages used in the study stage were assigned. Passage topics are also indicated.

Table 2

LanguEdge Test Forms and Text Topics

LanguEdge test form	Subjects assigned test for pretest[a, b]	Text [c]	Subjects assigned test/text for study[d]	Text topic
		1	C3, C5, C8; J2, J4, J8; K3, K6; O1, O6, O7	The expression of emotions
1	C1, C2, C4, C6; J1, J3, J6, J7; K2, K5, K7, K8; O3, O4, O5, O8	2	C3, C7, C8; J4, J5; K1, K3, K4; O1, O2, O6	Nineteenth-century politics in the U.S.
		3	C5, C7; J2, J5, J8; K1, K4, K6; O2, O7	Geology and landscape Desert formation
2	C3, C5, C7, C8; J2, J4, J5, J8; K1, K3, K4, K6; O1, O2, O6, O7	2	C1, C4, C6; J1, J3, J6; K2, K7; O4, O5	Early cinema
		3	C1, C2, C4; J3, J7; K5, K7, K8; O3, O5, O8	The origins of Cetaceans

[a]C, J, K, O indicate Chinese, Japanese, Korean, and Other subjects, respectively; C1 = Chinese subject 1, etc. [b]Subjects completed all questions on all three texts for the test form assigned in the pretest. [c]Both tests include three reading texts (passages), with each text followed by 12–13 test items. [d]Subjects completed all questions for only two of the three texts, as indicated, for the test form assigned for the study.

Description of Reading Test Item Types

As stated previously, the new TOEFL reading section uses three general item types to evaluate the reader's proficiency with regards to accomplishing typical academic-like reading tasks, specifically: Basic Comprehension items, Inferencing items, and Reading to Learn items. ETS has defined five different types of Basic Comprehension items, three different types of Inferencing items, and two different types of Reading to Learn items, for a total of ten different item types. Appendix E provides a brief definition and description of each of these item types along with an example of how each item type is operationalized in the new TOEFL.

Data Collection Sessions

For the pilot study, three sessions of approximately two hours each were scheduled. The first session was used to complete the consent form, background questionnaire, LanguEdge training, and LanguEdge reading pretest to determine reading proficiency. The second session was used to provide training on giving verbal reports, and then to complete the two sets of the study form of the LanguEdge reading test. The third session was used to complete the last set from the LanguEdge and to conduct a retrospective interview. It was determined that this three-session structure was undesirable for three primary reasons. First, it proved logistically challenging to schedule subjects for three separate sessions; second, the sheer amount of video collected made it nearly impossible to review and prepare it in a timely manner to meaningfully use it as part of a retrospective interview during the final session; and third, the amount of data collected would clearly be too great to transcribe/translate and code in a timely manner for analysis, especially since the verbal reports could go several hours in length.

Consequently, for the main study, the following modifications were made to the data collection sessions. First, the number of sessions was reduced from three to two. Second, subjects were asked to respond to and provide a verbal report on just two of the three reading sets on the LanguEdge courseware. Third, retrospective interviews were conducted immediately following the completion of each test set. The agenda for the two sessions follows:

Session #1 included: (~2 hours)

- Completion of the informed consent form (~10 minutes)
- Background questionnaire (~5 minutes)

- Training in responding to the LanguEdge (computer-based; ~20 minutes)
- LanguEdge reading pretest (~75 minutes)
- Oral report of strategies on the last test item (~5 minutes)

Session #2 included: (~3.5 hours)

- Respondent training in giving verbal report (~30 minutes)
- Completion of two passages from LanguEdge reading test while doing concurrent and retrospective verbal report (~ 3 hours)

Furthermore, only representative samples of every item type from each of the reading sets were translated/transcribed and coded, although subjects completed and provided a verbal report on all the items in the two sets they did. This is described in more detail in the Data Analysis section of this report.

Research Assistants

In order to facilitate the collection of native-language data from our respondents, the RAs for the study were native speakers of the respondents' languages, except in the case of four respondents (one speaker each of Bengali and Thai and two of Arabic). All four RAs were doctoral students at the University of Minnesota, and all had studied research methods in their course work. The RAs were provided with an initial orientation, training at the pilot phase, and then regular guidance through the data collection and data analysis phases (see Table 3).

Table 3

Research Assistants' Background

Role	First language	M/F	Level and field of study
Principal RA	Turkish	M	Ph.D.—Human resources development
RA	Chinese	M	Ph.D.—Cross-cultural education
RA	Japanese	F	Ph.D.—Cross-cultural education
RA	Korean	F	Ph.D.—Cross-cultural education

Data Analysis

Item Selection

Due to the sheer volume of verbal report data collected (an average of over three hours of digital video per subject), cost and time constraints mandated that only a subset of the data be analyzed because of the time-intensive nature of the analysis process.[6] Hence, prior to the start of the data collection, it was determined that the verbal reports for each respondent on 13 predetermined items across the two completed reading test sets would be transcribed/translated and analyzed for strategy use. Distribution of the items to be analyzed was determined by considering a variety of factors, including the following:

1. Each general item type (Basic Comprehension, Inferencing, Reading to Learn) should be represented in both of the reading test sets completed by each subject.
2. When possible, each of the 10 item subtypes (such as Basic Comprehension-vocabulary and Inferencing-rhetorical purpose) should be represented in at least one if not both of the reading test sets completed by each subject.
3. All questions should be used at approximately the same rate as the other questions of the same item type across the six reading test sets.
4. Verbal report data should be collected from respondents (Chinese, Japanese, Korean, Other) as they are answering all the questions across all six reading test sets.

The majority of the item types across the six test sets were categorized by the TOEFL test construction committee as assessing vocabulary and factual information within the Basic Comprehension category.[7] Consequently, two of both of these item types, one from each of the two test sets completed, were analyzed for each subject. One instance of each of the three remaining Basic Comprehension item types (pronoun reference, sentence simplification, and negative fact) was selected from across the two test sets completed by each subject, while addressing, if possible, the factors listed above. This meant that a total of seven Basic Comprehension questions were selected for analysis per subject.

Four Inferencing item types were analyzed for each subject, two from each reading set completed. However, since the distribution of the three Inferencing item types (I, I-it, I-rp) was less consistent and equitable across the six reading test sets in the LanguEdge materials, some of the Inferencing item types were not analyzed for all 32 subjects.

Each of the six reading test sets included only one Reading to Learn item type, with five of the six calling for a R2L-prose summary. Only one reading test set used the R2L-schematic table item type. Both of the R2L items completed by each subject were included in the analysis.

The general plan for the selection of 13 test items to be analyzed for each subject is given in Table 4. The subject-by-subject distribution of the test items that were analyzed is provided in Table 5. Table 6 lists the items by type, and further ordered by test, passage, and question.

Table 4

General Distribution of Test Items Analyzed Per Subject

Test item type	Reading 1	Reading 2	Reading 1 or 2
BC-f	1	1	
BC-v	1	1	
BC-ss; BC-n/e; BC-pr[a]			3
I; I-it; I-rp[b]	2	2	
R2L-pr; R2L-st[c]	1	1	

Note. There were a couple of minor exceptions to this general distribution plan because some readings did not include certain test item types.

[a]One item representing each of these three Basic Comprehension (BC) types was chosen when possible from across the two readings. [b]At least one item representing each of the Inferencing (I) types was chosen when possible—two from each reading. [c]Each reading had only one R2L item type.

Table 5

Item Type Analyzed by Subject and Test Item Number

Subject	BC-v	BC-f	BC-ss	BC-n/e	BC-pr	I	I-it	I-rp	R2L-ps	R2L-st
J1	T2P1Q3 T2P2Q11	T2P1Q6 T2P2Q8	T2P1Q10	T2P2Q1	T2P2Q9	T2P1Q11	T2P1Q12 T2P2Q12	T2P2Q2	T2P1Q13 T2P2Q13	
J2	T1P1Q9 T1P3Q2	T1P1Q5 T1P3Q5	T1P1Q6 T1P3Q9		T1P3/Q8	T1P3Q3	T1P1Q12 T1P3Q11	T1P1Q2	T1P1Q13	T1P3Q12
J3	T2P2Q4 T2P3Q3	T2P2Q10 T2P3Q1	T2P2Q3	T2P2Q1	T2P3Q5	T2P3Q9	T2P2Q12 T2P3Q12	T2P2Q2	T2P2Q13 T2P3Q13	
J4	T1P1Q10 T1P2Q8	T1P1Q7 T1P2Q3	T1P1Q6	T1P2Q10	T1P1Q4	T1P2Q9	T1P1Q12 T1P2Q12	T1P1Q2	T1P1Q13 T1P2Q13	
J5	T1P2Q1 T1P3Q2	T1P2Q2 T1P3Q5	T1P2Q11	T1P2Q10	T1P3Q8	T1P2Q9	T1P2Q12 T1P3Q11	T1P3Q6	T1P2Q13	T1P3Q12
J6	T2P1Q5 T2P2Q11	T2P1Q4 T2P2Q10	T2P2Q3	T2P1Q9	T2P2Q9	T2P1Q11	T2P1Q12 T2P2Q12	T2P2Q2	T2P1Q13 T2P2Q13	
J7	T2P1Q1 T2P3Q11	T2P1Q8 T2P3Q4	T2P3Q10	T2P1Q9	T2P3Q5	T2P1Q11 T2P3Q8	T2P1Q12 T2P3Q12		T2P1Q13 T2P3Q13	
J8	T1P1Q1 T1P3Q7	T1P1Q8 T1P3Q10	T1P1Q6 T1P3Q9		T1P1Q4	T1P3Q3	T1P1Q12 T1P3Q11	T1P1Q2	T1P1Q13	T1P3Q12
C1	T2P2Q4 T2P3Q6	T2P2Q7 T2P3Q4	T2P2Q3 T2P2Q10		T2P3Q5	T2P3Q2 T2P3Q9	T2P2Q12	T2P2Q2	T2P2Q13 T2P3Q13	
C2	T2P1Q1 T2P3Q11	T2P1Q2 T2P3Q7	T2P1Q10	T2P1Q9	T2P3Q5	T2P1Q11 T2P3Q8	T2P1Q12 T2P3Q12		T2P1Q13 T2P3Q13	
C3	T1P1Q1 T1P2Q8	T1P1Q7 T1P2Q2	T1P2Q11	T1P2Q10	T1P1Q4	T1P2Q9	T1P1Q12 T1P2Q12	T1P1Q2	T1P1Q13 T1P2Q13	

(Table continues)

Table 5 (continued)

Subject	BC-v	BC-f	BC-ss	BC-n/e	BC-pr	I	I-it	I-rp	R2L-ps	R2L-st
C4	T2P2Q5 T2P3Q3	T2P2Q6 T2P3Q7	T2P3Q10	T2P2Q1	T2P3Q5	T2P3Q2 T2P3Q9	T2P2Q12	T2P2Q2	T2P2Q13 T2P3Q13	
C5	T1P1Q9 T1P3Q4	T1P1Q8 T1P3Q1	T1P3Q9		T1P1Q4 T1P3Q8	T1P3Q3	T1P1Q12 T1P3Q11	T1P1Q2	T1P1Q13	T1P3Q12
C6	T2P1Q7 T2P2Q11	T2P1Q4 T2P2Q8	T2P1Q10	T2P1Q9	T2P2Q9	T2P1Q11	T2P1Q12 T2P2Q12	T2P2Q2	T2P1Q13 T2P2Q13	
C7	T1P2Q1 T1P3Q2	T1P2Q7 T1P3Q5	T1P3Q9	T1P2Q10	T1P3Q8	T1P2Q9	T1P2Q12 T1P3Q11	T1P3Q6	T1P2Q13	T1P3Q12
C8	T1P1Q3 T1P2Q6	T1P1Q7 T1P2Q3	T1P1Q6	T1P2Q10	T1P1Q4	T1P2Q9	T1P1Q12 T1P2Q12	T1P1Q2	T1P1Q13 T1P2Q13	
K1	T1P2Q8 T1P3Q4	TIP2Q3 T1P3Q1	T1P3Q9	T1P2Q10	T1P3Q8	T1P2Q9 T1P3Q3	T1P2Q12	T1P3Q6	T1P2Q13	T1P3Q12
K2	T2P1Q5 T2P2Q5	T2P1Q4 T2P2Q6	T2P2Q3	T2P1Q9	T2P2Q9	T2P1Q11	T2P1Q12 T2P2Q12	T2P2Q2	T2P1Q13 T2P2Q13	
K3	T1P1Q3 T1P2Q6	T1P1Q11 T1P2Q4	T1P1Q6	T1P2Q10	T1P1Q4	T1P2Q9	T1P1Q12 T1P2Q12	T1P1Q2	T1P1Q13 T1P2Q13	
K4	T1P2Q1 T1P3Q7	T1P2Q4 T1P3Q10	T1P2Q11	T1P2Q10	T1P3Q8	T1P2Q9 T1P3Q3	T1P2Q12	T1P3Q6	T1P2Q13	T1P3Q12
K5	T2P1Q3 T2P3Q3	T2P1Q6 T2P3Q1	T2P1Q10	T2P1Q9	T2P3Q5	T2P1Q11 T2P3Q8	T2P1Q12 T2P3Q12		T2P1Q13 T2P3Q13	
K6	T1P1Q10 T1P3Q4	T1P1Q5 T1P3Q1	T1P1Q6 T1P3Q9		T1P1Q4	T1P3Q3	T1P1Q12 T1P3Q11	T1P1Q2	T1P1Q13	T1P3Q12

(Table continues)

Table 5 (continued)

Subject	BC-v	BC-f	BC-ss	BC-n/e	BC-pr	I	I-it	I-rp	R2L-ps	R2L-st
K7	T2P2Q4 T2P3Q6	T2P2Q7 T2P3Q7	T2P3Q10	T2P2Q1	T2P3Q5	T2P3Q2 T2P3Q9	T2P2Q12	T2P2Q2	T2P2Q13 T2P3Q13	
K8	T2P1Q7 T2P3Q11	T2P1Q2 T2P3Q4	T2P1Q10	T2P1Q9	T2P3Q5	T2P1Q11 T2P3Q8	T2P1Q12 T2P3Q12		T2P1Q13 T2P3Q13	
O1	T1P1Q3 T1P2Q1	T1P1Q8 T1P2Q5	T1P1Q6	T1P2Q10	T1P1Q4	T1P2Q9	T1P1Q12 T1P2Q12	T1P1Q2	T1P1Q13 T1P2Q13	
O2	T1P2Q6 T1P3Q7	T1P2Q7 T1P3Q10	T1P2Q11	T1P2Q10	T1P3Q8	T1P2Q9	T1P2Q12 T1P3Q11	T1P3Q6	T1P2Q13	T1P3Q12
O3	T2P1Q7 T2P3Q6	T2P1Q8 T2P3Q4	T2P3Q10	T2P1Q9	T2P3Q5	T2P1Q11 T2P3Q2 T2P3Q8			T2P1Q13 T2P3Q13	
O4	T2P1Q1 T2P2Q11	T2P1Q2 T2P2Q8	T2P1Q10	T2P2Q1	T2P2Q9	T2P1Q11	T2P1Q12 T2P2Q12	T2P2Q2	T2P1Q13 T2P2Q13	
O5	T2P2Q4 T2P3Q11	T2P2Q7 T2P3Q1	T2P3Q10	T2P2Q1	T2P2Q9	T2P3Q2 T2P3Q9	T2P2Q12	T2P2Q2	T2P2Q13 T2P3Q13	
O6	T1P1Q1 T1P2Q8	T1P1Q11 T1P2Q4	T1P2Q11	T1P2Q10	T1P1Q4	T1P2Q9	T1P1Q12 T1P2Q12	T1P1Q2	T1P1Q13 T1P2Q13	
O7	T1P1Q9 T1P3Q4	T1P1Q7 T1P3Q5	T1P1Q6 T1P3Q9		T1P3Q8	T1P3Q3	T1P1Q12	T1P1Q2 T1P3Q6	T1P1Q13	T1P3Q12
O8	T2P1Q3 T2P3Q3	T2P1Q6 T2P3Q7	T2P1Q10	T2P1Q9	T2P3Q5	T2P1Q11 T2P3Q8	T2P1Q12 T2P3Q12		T2P1Q13 T2P3Q13	

Note. For test items, T = LanguEdge test (1 or 2), P = test set (passage) in LanguEdge test (1, 2, or 3), Q = item (question) in test set (1–13). See the Test Materials section in this report for a description of item type abbreviations (e.g., BC-v = item type Basic Comprehension-vocabulary).

Table 6

Item Type Distribution by Set, Item Number, and Subject

Item type	Test set	Item # (Q)	Subject
BC-vocabulary (BC-v)	T1P1	1	C3, O6, J8
		3	O1, K3, C8
		9	J2, C5, O7
		10	J4, K6
	T1P2	1	O1, K4, J5, C7
		6	O2, K3, C8
		8	K1, C3, J4, O6
	T1P3	2	J2, J5, C7
		4	K1, C5, O7, K6
		7	O2, K4, J8
	T2P1	1	C2, O4, J7
		3	J1, K5, O8
		5	K2, J6
		7	O3, C6, K8
	T2P2	4	C1, J3, O5, K7
		5	K2, C4
		11	J1, O4, J6, C6
	T2P3	3	J3, C4, K5, O8
		6	C1, O3, K7
		11	C2, O5, J7, K8
BC-fact (BC-f)	T1P1	5	J2, K6
		7	C3, J4, O7, C8
		8	O1, C5, J8
		11	K3, O6
	T1P2	2	C3, J5
		3	K1, J4, C8
		4	K3, K4, O6

(Table continues)

Table 6 (continued)

Item type	Test set	Item # (Q)	Subject
		5	O1
		7	O2, C7
	T1P3	1	K1, C5, K6
		5	J2, J5, O7, C7
		10	O2, K4, J8
	T2P1	2	C2, O4, K8
		4	K2, J6,C6
		6	J1, K5, O8
		8	O3, J7
	T2P2	6	K2, C4
		7	C1, O5, K7
		8	J1, O4, C6
		10	J3, J6
	T2P3	1	J3, O5, K5
		4	C1, O3, J7, K8
		7	C2, C4, K7, O8
BC-other	T1P1		
not/except (BC-n/e)		--	
sentence simplification (BC-ss)		6	O1, J2, K3, J4, K6, O7, J8, C8
pronoun reference (BC-pr)		4	O1, C3, K3, J4, C5, K6, O6, J8, C8
	T1P2		
not/except BC-n/e)		10	K1, O1, O2, C3, K3, K4, J4, J5, O6, C7, C8
sentence simplification (BC-ss)		11	O2, C3, K4, J5, O6
pronoun reference (BC-pr)		--	

(Table continues)

Table 6 (continued)

Item type	Test set	Item # (Q)	Subject
	T1P3		
not/except (BC-n/e)		--	
sentence simplification (BC-ss)		9	K1, J2, C5, C7, O7, J8, K6
pronoun reference (BC-pr)		8	K1, J2, O2, K4, C5, J5, C7, O7
	T2P1		
not/except (BC-n/e)		9	C2, K2, O3, K5, J6, C6, J7, K8, O8
sentence simplification (BC-ss)		10	J1, C2, O4, K5, C6, K8, O8
pronoun reference (BC-pr)		--	
	T2P2		
not/except (BC-n/e)		1	J1, J3, O4, C4, O5, K7
sentence simplification (BC-ss)		3	C1, K2, J3, J6
pronoun reference (BC-pr)		9	J1, K2, O4, O5, J6, C6
	T2P3		
not/except (BC-n/e)		--	
sentence simplification (BC-ss)		10	C1, O3, C4, O5, K7, J7
pronoun reference (BC-pr)		5	C1, C2, J3, O3, C4, K5, J7, K8, O8
I (Inferencing)	T1P1		
inference (I)		--	
insert text (I-it)		12	O1, J2, C3, K3, J4, C5, O6, O7, J8, C8, K6
rhetorical purpose (I-rp)		2	O1, J2, C3, K3, J4, C5, O6, O7, J8, C8, K6
	T1P2		
inference (I)		9	K1, O1, O2, C3, K3, K4, J4, J5, O6, C7, C8

(Table continues)

Table 6 (continued)

Item type	Test set	Item # (Q)	Subject
insert text (I-it)		12	K1, O1, O2, C3, K3, K4, J4, J5, O6, C7, C8
rhetorical purpose (I-rp)		--	
	T1P3		
inference (I)		3	K1, J2, K4, C5, O7, J8, K6
insert text (I-it)		11	J2, O2, C5, J5, C7, J8, K6
rhetorical purpose (I-rp)		6	K1, O2, K4, J5, C7, O7
	T2P1		
inference (I)		11	J1, C2, K2, O3, O4, K5, J6, C6, J7, K8, O8
insert text (I-it)		12	J1, C2, K2, O3, O4, K5, J6, C6, J7, K8, O8
rhetorical purpose (I-rp)		--	
	T2P2		
inference (I)		--	
insert text (I-it)		12	J1, C1, K2, J3, O4, C4, O5, J6, K7
rhetorical purpose (I-rp)		2	J1, C1, K2,J3, O4, C4, O5, J6, K7
	T2P3		
inference (I)		2	C1, O3, C4, O5, K7
		8	C2, O3, K5, J7, K8, O8
		9	C1, J3, C4, O5, K7
insert text (I-it)		12	C2, J3, K5, J7, K8, O8
rhetorical purpose (I-rp)		--	

(Table continues)

Table 6 (continued)

Item type	Test set	Item # (Q)	Subject
R2L (Reading to Learn)			
prose summary (R2L-pr)	T1P1	13	J2, J4, J8, C3, C5, C8, K3, K6, O1, O6, O7
	T1P2	13	J4, J5, C3, C7, C8, K1, K3, K4, O1, O2, O6
	T2P1	13	J1, J6, J7, C2, C6, K2, K5, K8, O3, O4, O8
	T2P2	13	J1, J3, J6, C1, C4, C6, K2, K7, O4, O5
	T2P3	13	J3, J7, C1, C2, C4, K5, K7, K8, O3, O5, O8
schematic table (R2L-st)	T1P3	12	J2, J5, J8, C5, C7, K1, K4, K6, O2, O7

Note. Test set symbols indicate which test and which test passage taken on the LanguEdge courseware materials (e.g., T1P1 = test one, passage one). Item # (Q) indicates the item (i.e., question) in the test passage.

Strategy Coding

The researchers and research assistants, drawing on the literature reviewed above on reading strategies and test-taking strategies, developed preliminary rubrics for reading and test-taking strategies (test management and test-wiseness) that were used to code the verbal reports produced during the pilot study. A series of lengthy working sessions involving the researchers and the four RAs provided important input for the development of the final set of rubrics for coding the strategies revealed by means of verbal report in the main study. The set of codes was reviewed and discussed extensively a second time early in the coding phase of they study, with minor modifications made to improve clarity. These codes were further modified or combined once all the data was collected and analyzed by the researchers to better reflect the strategies actually used by the respondents. The final version of the rubrics used for the reading, test-management, and test-wiseness strategies in the analysis of the verbal reports can be found in Tables 7, 8, and 9.

Table 7

Revised Reading Strategies Coding Rubric (R)

Strategy	Description
Approaches to reading the passage	
R1	Plans a goal for the passage.
R2	Makes a mental note of what is learned from the prereading.
R3	Considers prior knowledge of the topic.
R4	Reads the whole passage carefully.
R5	Reads the whole passage rapidly.
R6	Reads a portion of the passage carefully.
R7	Reads a portion of the passage rapidly looking for specific information .
R8	Looks for markers of meaning in the passage (e.g., definitions, examples, indicators of key ideas, guides to paragraph development).
R9	Repeats, paraphrases, or translates words, phrases, or sentences—or summarizes paragraphs/passage—to aid or improve understanding.
R10	Identifies an unknown word or phrase.
R11	Identifies unknown sentence meaning.
Uses of the passage and the main ideas to help in understanding	
R12	During reading rereads to clarify the idea.
R13	During reading asks self about the overall meaning of the passage/portion.
R14	During reading monitors understanding of the passage/portion's discourse structure (e.g., compare/contrast, description, definition).
R15	Adjusts comprehension of the passage as more is read: Asks if previous understanding is still accurate given new information.
R16	Adjusts comprehension of the passage as more is read: Identifies the specific new information that does or does not support previous understanding.
R17	Confirms final understanding of the passage based on the content and/or the discourse structure.

(Table continues)

Table 7 (continued)

Strategy	Description
Identification of important information and the discourse structure of the passage	
R18	Uses terms already known in building an understanding of new terms.
R19	Identifies and learns the key words of the passage.
R20	Looks for sentences that convey the main ideas.
R21	Uses knowledge of the passage/portion: Notes the discourse structure of the passage /portion (cause/effect, compare/contrast, etc.).
R22	Uses knowledge of the passage/portion: Notes the different parts of the passage (introduction, examples, transitions, etc.) and how they interrelate ("Is this still part of the introduction or is this the first topic?" "This sounds like a summary—is it the conclusion?").
R23	Uses knowledge of the passage/portion: Uses logical connectors to clarify content and passage organization (e.g., "First of all," "On the other hand," "In conclusion").
R24	Uses other parts of the passage to help in understanding a given portion: Reads ahead to look for information that will help in understanding what has already been read.
R25	Uses other parts of the passage to help in understanding a given portion: Goes back in the passage to review/understand information that may be important to the remaining passage.
Inferences	
R26	Verifies the referent of a pronoun.
R27	Infers the meanings of new words by using work attack skills: Internal (root words, prefixes, etc.).
R28	Infers the meanings of new words by using work attack skills: External context (neighboring words/sentences/overall passage).

Table 8

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.

(Table continues)

Table 8 (continued)

Strategy	Description
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 item type).
T23	Selects options through elimination of other option(s) as unreasonable based on background knowledge.
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)h 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.

Table 9

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.

While the revised coding for most of the strategies is relatively straightforward, it is necessary to explain why "repeats, paraphrases, or translates words, phrases, or sentences" is combined with "summarizes paragraphs/passage" as reading strategy R9 (Table 7). These were originally two separate strategies, but were combined based on empirical evidence of actual strategy use. First, almost all instances of summarizing occurred at the paragraph and even sentence level whenever it occurred in the context of responding to the test items. In this respect, this was almost always a limited text strategy. Subjects for the most part simply did not try to summarize the whole passage on their own. Because any summarizing was almost always done with only shorter chunks of texts, the raters had a very difficult time distinguishing in practice between a summary and a paraphrase. There were no instances of subjects trying to summarize the whole passage before or while trying to answer the one item type that might have been expected to prompt that strategy, namely, the R2L items. As will be noted in the Results section of this report, subjects focused on the items as a test-taking task and so their primary goal was to get the item right. Because they already had a strong understanding of the text from working through the previous 12 test items that always come before the R2L item, they typically had no need to take the time to read and summarize the passage in their heads—even if that was the intent of the R2L item design.

The process for identifying the strategies used by subjects was approached in much the same way that *moves* in a discourse genre are identified. A genre move refers to a section of a text that performs a specific communicative function; each move not only has its own purpose but also contributes to the overall communicative purpose of the genre (Bhatia, 1993; Swales, 1981, 1990). While a genre move represents a recognizable communicative event characterized by a communicative purpose, a reading or test-taking strategy represents a specific and recognizable strategic choice made by the subject that is deliberate and purposeful and is intended to facilitate the reading or test-taking task. Furthermore, just as a genre move is identified and mutually understood by members of the professional or academic community in which it regularly occurs, a reading or test-taking strategy can also be identified and mutually understood by expert readers and test-takers. Verbalizations were classified into different strategies by determining what strategic function they revealed, including explicitly stated as well as obviously implied strategies. Consequently, strategy units had no set length.

Most strategies tended to be represented by sentences, but some were signaled by single words, others by multiple sentences. It was also the case that multiple strategies (e.g., a reading and a test-taking strategy) could be signaled by the same verbalization. Accurately identifying the different strategies and segmenting the transcribed verbalizations was a high priority. After extensive group orientation sessions for ensuring consistency, the research assistants did the initial translation/transcription and strategy coding for each of the subjects for whom they collected data. All strategy coding was then checked by at least one other RA, with any discrepancies in coding addressed through discussion. One of the researchers then checked coding one final time during analysis to ensure accuracy.

Also with regard to strategy coding, the researchers did not include as strategies presumed strategies. In other words, only included were strategies that were consciously referred to, not those that we might have intuited to exist outside the level of respondents' consciousness. Moves that were unconscious were simply viewed as processes, not as strategies. So, for example, if the RAs found an instance of test-taking strategy T22, "Selects options through vocabulary, sentence, paragraph, or passage overall meaning," even if it were presumed that the respondent would have had to employ multiple reading strategies (from the R12-R28 list) to arrive at passage overall meaning, reading strategies were only coded if specifically reported by the subjects. In fact, the data suggest that these reading strategies were employed at some prior point in the test—either when the subjects first read the passage or while they worked through previous items. Furthermore, respondents by and large verbalized more their test-taking strategies than their reading strategies, although as noted many of the test-taking strategies involved reading as well. Also, many respondents only read the text as needed in order to answer the test items. They never really read through the passage from start to finish as a coherent text until the R2L question (if then) when they already would have had a good idea of what the passage was about from having worked through the other questions. This precluded their using reading strategies they might have otherwise employed.

Lastly, it should be noted that verbal reports were coded as either being concurrent (involving thinking aloud, introspection, or immediate retrospection) or involving delayed or prompted retrospection, which occurred after the respondents had completed the items on the given version of the test or as a result of a targeted prompt by the research assistants (e.g., where the RA would ask questions like the following at the end of a subject's verbalization on a specific

item: “Did you confirm your answer by replacing the idea in the related sentence?”). Only *concurrent* verbal reports rather than delayed retrospection through prompted elicitations were used in the frequency counts of strategy use as will be described. While the researchers did use the delayed retrospective verbal reports to help in understanding what the subjects were doing, the choice was to be conservative in not including the subjects’ responses following these kinds of prompts. The concern was to insure the integrity of the verbal reports by avoiding the use of verbal reports that were removed in time from the actual mental events. Furthermore, not all the RAs consistently included the same kind of delayed verbal elicitations in their final coded data. Nevertheless, there were not very many instances of delayed retrospection, and there was more than enough data without them.

Data Tagging

In order to facilitate the analysis of strategy occurrence and distribution across question types, a tagging system was developed to identify each of the strategy occurrences that had been coded in the verbal reports. This allowed for both the qualitative and quantitative analysis of the corpus using the software programs Microsoft Word and SPSS. Appendix F provides an overview of the tagging system that was used.

Procedures for Data Analysis

Quantitative analysis. While this study was by its very nature primarily a qualitative one, it was felt that some effort at quantifying the verbal report data would help to lend more rigor to statements about the frequency of reading and test-taking strategy use. Hence, the coding scheme was developed to count the occurrence of both reading and test-taking strategies in as finely tuned a manner as possible. As the coding proceeded, it became clear that some categories needed to be collapsed since the coders were not actually making such finely tuned distinctions. The principal variables in the study, such as strategy types and item types, were measured as nominal variables and the total number of times a strategy was verbalized or otherwise indicated was tallied. Consequently, it was possible that a single test question would prompt multiple instances of a particular strategy. The complexity of the resulting coding necessary to account for these multiple entries precluded the use of statistical measures typically run on nominal variables, such as chi-square.

Once all the individual occurrences of reading and test-taking strategies were identified, coded, tagged for SPSS data analysis, and analyzed, the challenge was to come up with some system for rigorously distinguishing levels of frequency of occurrence. Because there was variety in the total number of questions per item type sampled in this study across the six reading passages (e.g., there were 64 Basic Comprehension fact items and only 10 Reading to Learn schematic table items), a simple count of strategy frequencies would have distorted the importance of any given strategy. Hence, the raw strategy totals were converted into ratio scores using a type/token analysis: the ratio of number of occurrences of each strategy type in relation to the total number of items of a type used in data collection for the study. As noted above, in the interests of time and resources, subjects in the study completed only two of the three passages, and while they provided verbal reports on their responses for all the items, the verbal report data for only a balanced sampling of these items were analyzed (see Table 5).

The ratio scores derived from this analysis were then categorized by frequency in a partly empirical and partly intuitive way, relying more on qualitative criteria than on any hard and fast quantitative measure of intervals.[8] The reason was that the finest distinctions seemed to be at the lower levels of frequency, between low, moderate, and high frequency of strategy use. At the highest frequency level, the numbers tended to become inflated by multiple uses of the same strategy by the same individual because multiple counting was allowed. The cut off points used were as follows:

very high (VH) frequency	≥ 1.00
high (H) frequency	≥ 0.50
moderate (M) frequency	≥ 0.30
low (L) frequency	≤ 0.29

In other words, a type/token frequency of 1.32, for example, was rated as very high (VH), a type/token frequency of .68 was rated as high, a type/token ratio of .49 was rated as moderate, and a type/token frequency of .03 was rated as low.

In addition, this quantitative analysis does not reflect hard and fast numbers, but trends since not all strategies were verbalized or verbalized every time they are used. With regard to the

nonverbalized strategies, ones that were considered by the RAs to be obvious from the respondents' actions were coded.

Finally, the quantitative analysis also took into consideration the relationship that different strategies appeared to have with one another, regardless of the relative frequency with which these strategies were used. Likewise, strategies that might have been used only infrequently but that reflected noteworthy approaches to the reading and test-taking tasks were included in the analysis. In some instances, the analysis also included strategies that were seldom used, contrary to expectation.

Qualitative analysis. Once significant relationships were determined between item types, the specific patterns of strategy use were then more carefully examined, bearing in mind that strategies for reading and for test taking invariably cluster together in response patterns, sometimes in sequence and sometimes in groups. The analysis focused on the patterns of strategy use that best characterized the responses for each item type. The intention of the analysis was to produce a series of examples for each strategy that would help to provide a qualitative description of what the response process actually consisted of across the respondents. The analysis paid close attention to whether the reported processes for responding to a given item were consistent with the aims of the test constructors and hence indicated that the item was testing what it purported to test.

Results

As would be expected, the specific use of strategies by a given respondent was, to a certain extent, idiosyncratic in nature in that no two respondents approached any particular test item in exactly the same way. Nevertheless, due to the nature of the clear-cut task at hand, strategy trends did nonetheless emerge from the data, with respondents relying primarily on specific reading and test-taking strategies that can be associated with academic-like tasks involving the reading of passages and responding to the questions about them.

The sections Frequency of Strategy Use Across All Item Types and Detailed Analysis of Strategy Use by Item Type will deal with the answer to the first research question concerning the reading strategies and test-taking strategies that respondents reported using for completing each of the ten different test item types. The section Comparison of the Inferencing and Reading to Learn Items with the More Traditional Basic Comprehension Items will briefly respond to the issue of whether the Reading to Learn and the Inferencing items required and assessed somewhat different

approaches to academic reading from those associated with the Basic Comprehension questions. It will also address the issue of item difficulty.

Frequency of Strategy Use Across All Item Types

A summary table of the quantitative results (Table 10) provides the strategy use frequencies across item types. Then a summary of how to interpret the examples for each item type follows. The rest of the Results section consists of a report by item type where the frequencies are first presented and then followed up with examples, in order to depict in a more descriptive way the processes involved in responding to the given item type.

Table 10

Frequency of Reported Use of Reading and Test-Taking Strategies[9]

Strategy	BC-v	BC-pr	BC-ss	BC-f	BC-n/e	I	I-rp	I-it	R2L-ps	R2L-st
R6	H	VH	VH	VH	VH	VH	VH	VH	H	VH
R7	L	M	L	L	L	L	M	H	H	H
R9	M	H	H	VH	H	H	H	H	M	VH
R10	L	L	L	M	L	L	L	L	L	L
R26	L	H	L	L	L	L	L	L	L	L
T1	M	M	M	H	H	H	H	M	M	H
T2	L	L	M	H	M	H	M	M	M	VH
T3	L	L	L	L	L	M	L	L	L	L
T4	L	L	L	L	L	L	L	L	H	H
T5	VH	VH	VH	VH	VH	VH	VH	VH	L	L
T6	L	L	L	L	L	L	M	L	L	L
T8	L	L	L	L	L	L	L	M	L	L

(Table continues)

Table 10 (continued)

Strategy	BC-v	BC-pr	BC-ss	BC-f	BC-n/e	I	I-rp	I-it	R2L-ps	R2L-st
T10	H	M	L	L	L	L	L	L	L	L
T12	H	L	H	H	H	H	H	L	L	L
T13	H	L	L	L	L	L	L	L	L	L
T14	L	L	M	M	M	H	M	L	H	L
T16	H	H	VH	VH	VH	VH	VH	L	VH	VH
T17	L	L	M	L	L	L	M	L	M	H
T19	L	L	M	L	M	L	L	L	H	L
T21	H	L	L	L	L	L	L	L	L	L
T22	L	VH	H	H	H	H	H	H	VH	VH
T24	L	L	M	M	M	M	L	L	L	L
T26	L	L	L	L	L	L	L	M	L	L
T27	M	L	L	L	L	L	L	L	L	L
T28	VH	VH	VH	VH	VH	VH	VH	H	VH	VH

Note. Strategies that were used at a low (< .30) rate across *all* item types were not included in the table. Frequency rate = number of occurrences / number of items of that type. Rates ≥ 1.0 (marked VH) were classified as very high rates ≥ .50 (marked H) were classified as high rates ≥ .30 (marked M) were classified as moderate, and rates < .30 (marked L) were classified as low.

The strategy profiles for each of the item types, as reflected by the frequency rates indicated in Table 10, are discussed in Figure 1. The keys in Figure 1 should be referred to when interpreting the different uses of font styles and abbreviations used in the descriptions and examples.

KEY 1: Font Styles Used in Strategy Profiles and Examples

Plain text	In the examples, represents verbal reports in English (the L2).
Bold text	In the examples, represents verbal reports in the subjects' L1.
Italic text	Represents restatement of strategy definition in the strategy profiles as well as parenthetical observations about subject behavior in the examples when placed inside brackets, e.g., [*italic text is researcher note/observation of subject behavior*].
Underlined text	In the examples, represents text from the test passage.

KEY 2: Abbreviations Used in Strategy Profiles and Examples

R#	Reading Strategy (e.g., R1 = Reading Strategy 1. See Table 7.)
T#	Test-Management Strategy (e.g., T1 = Test-Management Strategy 1. See Table 8.)
TW#	Test-Wiseness Strategy (e.g., TW1 = Test-Wiseness Strategy 1. See Table 9.)
T#P#Q#	Specific Test Number, Passage Number, and Question Number of Example (e.g., T1P1Q1 = Test 1, Passage 1, Question 1 in the LanguEdge tests.)
J/C/K/O#	Specific subject number. J = Japanese subjects, C = Chinese subjects, K = Korean subjects, O = 'Other' subjects (e.g., J1 = Japanese subject #1. See Table 1.)
BC	Basic Comprehension item types. See Appendix E for description of the five different types: BC-v, BC-pr, BC-ss, BC-f, and BC-n/e.
I	Inferencing item types. See Appendix E for description of the three different types: I, I-rp, and I-it.
R2L	Reading to Learn item types. See Appendix E for a description of the two different types: R2L-ps and R2L-st.

Figure 1. **Keys to font styles and abbreviations used in item type descriptions and examples.**

Detailed Analysis of Strategy Use by Item Type

Basic Comprehension Items: Basic Comprehension-Vocabulary (BC-v)

The basic comprehension-vocabulary (BC-v) item type 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). Examinees need to select the option that can replace the targeted word while preserving the authors' intended meaning in the text context. (See Appendix E for an example of a BC-v item, as well as all the item types that will be described.) The accuracy rate for the 64 attempts to respond to the 20 different items of this type was 52/64 = 81% (J = 81%, C = 88%, K = 69%, O = 88%). Table 11 describes the most frequently used reading and test-taking strategies for this item type.

Table 11

Common Strategies for Item Type BC-v

Strategy code	Strategy description	Frequency rate
T5	Reads the question then reads the passage/portion to look for clues to the answer, either before or while considering options.	1.50
T28	Discards option(s) based on vocabulary, sentence, paragraph, or passage overall meaning as well as discourse structure (depending on item type).	1.03
T16	Considers the options and postpones consideration of the option.	.75
T10	Considers the options and checks the vocabulary option in context.	.69
R6	Reads a portion of the passage carefully.	.67
T13	Considers the options and defines the vocabulary option.	.50
T21	Selects options through background knowledge.	.50

(Table continues)

Table 11 (continued)

Strategy code	Strategy description	Frequency rate
R9	Repeats, paraphrases, or translates words, phrases, or sentences—or summarizes paragraphs/passage—to aid or improve understanding.	.48
T22	Selects options through vocabulary, sentence, paragraph, or passage overall meaning (depending on item type).	.45
T27	Discards option(s) based on background knowledge.	.39
T1	Goes back to the question for clarification: Rereads the question.	.33
T12	Considers the options and selects preliminary option(s) (lack of certainty indicated).	.25

Note. Strategies are ranked and grouped according to three levels of frequency rate (number of occurrences / number of items of that type): very high (≥ 1.0), high (≥ .50), and moderate (≥ .30).

In reviewing the reading and test-taking strategies that occurred at very high (≥ 1.0), high (≥ .50), and moderate (≥ .30) rates for this item type, as given in Table 11, the following notable strategy trends emerged:

1. Especially among those who did **not** recognize the word highlighted in the question, a strategy that occurred at a very high rate (1.50) was *jumping immediately to the word in the context of the passage before looking at the options to try to get a sense of the word's meaning* (T5). A reading strategy also occurred at a high rate (.67) along with this test-management strategy, namely, *reading a portion of the passage carefully* (R6). The following are examples of these two strategies, T5 and R6:

    - "[*Returns to passage*.] From sentence, 'devoid of' **means lack or not having**. Lack of, short of [*goes to the options*] hidden by [*Option B*], except for [*Option C*] lack in [*Option D*] yeah." (C6, T2P1Q7)
    - "**Well, this word** rate**.** [*Returns to passage*.] **Oh, when they report** positive feelings**, and** rate cartoons, **they become even happier.**" (C5, T1P1Q9)

- "'seep' seep? **I don't know this word. Let's go to the sentence in the text**." (K4, T1P3Q7)
- "**From the passage I could assume that it means** 'easily.'" (K7, T2P2Q4)
- "I don't know the meaning of 'propulsion.' So, I have to go and read the sentence in the passage." (O5, T2P3Q11)

2. Two of the commonly occurring strategy types naturally grouped together, namely, the respondents' use of their **background knowledge**[10] to make the final decision whether to select a particular option or to discard that option. These strategies were commonly used by subjects who already had a strong sense of the word for which they were being asked to find a synonym. *Using background knowledge to select* (T21) and *discard* (T27) options occurred at a high rate (.50) and a moderate rate (.39), respectively. This use of background knowledge to assist in the selection and discarding of options was most notable in the BC-v item type, with many subjects not even looking at the word in the context of the passage before choosing an option. Examples of strategies T21 and T27:

   - "Predictable **is not right . . .** Predictable **means predictable**." (J1, T2P1Q3)
   - "**The word** 'threatened' **should be defined as** '**making fearful.**' **So, I think it is natural to choose** 'endangered,' **which means** '**put in a dangerous situation**.'" (C2, T2P1Q1)
   - "Precious **means precious**. **So the answer is 'C**,' 'valuable.'" (J3, T2P3Q3)
   - "**The answer is** idea **since the word** concept **means** idea." (J4, T1P2Q8)

   Frequently occurring at the same time as the above strategies drawing on background knowledge, the strategy *defining vocabulary options as they are being considered* (T13) (to make sure they are comfortable with their meaning) also occurred at a high rate (.50). Examples focusing on strategy T13:

   - "'Exposed' in Chinese it means **discovered and made seen**." (C1, T2P3Q6)
   - "Contradictory means opposite, so this is not it." (K6, T1P1Q10)
   - [*Reads Option B.*] "Comparatively means **in comparison.** [*Reads Option C.*] Occasionally means **occasionally**. [*Reads Option D.*] Naturally means **naturally**." (J5, T1P3Q2)

3. Two other strategies that seemed to group together naturally were *using the understanding of the sentence and paragraph meaning to help select which option was the correct synonym* (T22) or *to discard options that weren't* (T28), which occurred at a moderate (.45) and very high rate (1.03), respectively. These two strategies also tended to co-occur with the strategies described in (1) above. Examples of strategies T22 and T28:

    - **"In this context the most likely answer should be B or D. Based on what I understand,** 'concur,' **it should be** 'agree' **rather than** 'understand.' **So, I choose B instead of D.**" (C8, T1P1Q3)
    - "Either dependent or applicable…[*reading passage*] to the British expression keep a stiff upper lip . . . is 'applicable' to this . . . so not 'dependent' . . . applicable is the right one." (K6, T1P1Q10)
    - **"I am sure that** 'obviously' **doesn't make sense. It's either** 'easily' **or** 'intelligently.' **For sure not** 'frequently'…**I think it's** 'easily' **because it's something about the effectiveness of the machine.** 'Easily' **makes more sense in the passage.**" (K7, T2P2Q4)

4. In keeping with the strategies noted in points (2) and (3) above that drew on background knowledge and context to choose options, the *consideration of options in context before a final decision is made* (T10) occurred at a high rate (0.69), with subjects often checking out the preferred option, or even all the options, in the context of the sentence before making a final decision. This strategy was by far more common in this item type than in any of the others. Examples of strategy T10:

    - "Let me read the passage now. 'Joy and sadness . . . despondent.' OK. This is the word that I didn't understand. So, I'm going to use the options to find out what it means. [*Reads Option A*]. OK. Joy and sadness . . . happy or 'curious'? I thought it would be the opposite of happy. So, I'm going to move to the other option. [*Reads Option B.*] OK. Joy and sadness . . . happy or unhappy? This might work because it's the opposite of each other. The remaining options are not opposites of happy. So, Option B is the correct answer." (O6, T1P1Q1)

- "There's another word 'do more easily, do easily.' I'm not sure here it is about time or frequency or the difficulty. So, I think maybe I need to read it in the context." (C1, T2P2Q4)
- "**Adverb. So I should choose an adverb. But the choices are all adverbs. I should insert all words**." (J3, T2P2Q4)
- "'Estimate' **does not make sense there**. 'Expect' **does not make sense there either**." (K3, T1P1Q3)
- "**I think Option A is not it because it does not make that much sense in context when I put it in the sentence. B is . . . Let's put B in context. I think B is the answer because it makes the most appropriate sense in context. And C and D are not it because I don't find their relevance to the weather; cold or warm**." (K4, T1P3Q7)
- "'Now located,' 'now visible.' Because if I put the word in here, I think 'located' will be a better choice: 'Located in the Sahara desert.'" (C1, T1P3Q6)

5. Reflecting the challenge of this item type for some respondents and/or the care they took in selecting their answer, *postponing the decision to choose or discard options until having reviewed the meanings of the options* (T16) and *making a preliminary (but uncertain) selection of an option* (T12) both occurred at a high rate (.75 and .52, respectively). These were frequently used along with strategy T10 described above. Examples of strategies T12 and T16:

    - "**The meaning of** expanded **is . . .** [*Reads Option A, then Option B, then Option C, then Option D.*] **I think the meaning of** expanded **is 'expand.' I think A is correct**." (J1, T2P2Q11)
    - "'Identified' means tell from; so that means has intention to figure out. 'Located' means landed." (C1, T1P3Q6)
    - "So it [*progressively*] means it has 'positively' grown. Positively? Progress positively? Uh, openly is not because it is not related to positive. Impressively is not positively. Objectively? No, it's not positively. Increasingly? Yes, because it is means 'positive.' Let me go back and check." (K2, T2P1Q5)

- "[*Reads Option D.*] **Maybe the answer is this one. But** [Option C] dependent to. Dependent to. Dependent to **sounds funny.**" (J4, T1P1Q10)
- "**I think the answer is B or D.** [*Rereads sentence in passage*.] The cultivation … **cultivation** … of crops has expanded … **expanded**… progressively drier regions…densities … **densities** … have grown. **The answer is either** B. impressively **or** D. increasingly." (J6, T2P1Q5)

6. There was only a low rate of occurrence of picking an option mainly because the other options were not viable (T23: .02, T11d: .22), and guessing an option because the others just didn't seem right (TW1: .09). Examples of strategies T23 and TW1:

    - "**I don't know the meaning of** 'relics' **but I think A is the answer…I know the meanings of C and D but not sure about B. B is confusing.**" (K1, T1P3Q4)
    - "**Then, I am going to put them one by one in context. A . . . B . . . C . . . D . . . this reads better in the sentence. I don't know exactly why but it** [*D*] **reads best.**" (K8, T2P3Q11)

Basic Comprehension-Pronoun Reference (BC-pr)

The basic comprehension-pronoun reference (BC-pr) item type 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). Examinees need to select the option that contains the noun or phrase that is referenced by the targeted pronoun. The accuracy rate for the 33 attempts to respond to the 4 items of this type was 29/33 = 88% (J = 88%, C = 89%, K = 88%, O = 88%). Table 12 outlines the most frequently used reading and test-taking strategies for this item type.

Table 12

Common Strategies for Item Type BC-pr

Strategy code	Strategy description	Frequency rate
T5	Reads the question then reads the passage/portion to look for clues to the answer, either before or while considering options.	2.44
R6	Reads a portion of the passage carefully.	1.67
T22	Selects options through vocabulary, sentence, paragraph, or passage overall meaning (depending on item type).	1.13
T28	Discards option(s) based on vocabulary, sentence, paragraph, or passage overall meaning as well as discourse structure (depending on item type).	.97
R9	Repeats, paraphrases, or translates words, phrases, or sentences—or summarizes paragraphs/passage—to aid or improve understanding.	.78
T16	Considers the options and postpones consideration of the option.	.59
T21	Selects options through background knowledge.	.50
R7	Reads a portion of the passage rapidly looking for specific information	.48
T10	Considers the options and checks the vocabulary option in context.	.38
R26	Verifies the referent of a pronoun.	.34
T1	Goes back to the question for clarification: Rereads the question.	.31

Note. Strategies are ranked and grouped according to three levels of frequency rate (number of occurrences / number of items of that type): very high (≥ 1.0), high (≥ .50), and moderate (≥ .03).

In reviewing the reading and test-taking strategies that occurred at very high (≥ 1.0), high (≥ .50), and moderate (≥ .30) rates for this item type, as given in Table 12, the following notable strategy trends emerge:

1. The most common strategy used by subjects was to *go directly to the passage (before looking at the options) and read the sentence that included the pronoun in question as well as the sentence or two before it* (T5), which occurred a very high rate (2.44). Three reading strategies that can be naturally grouped with this test-taking strategy include *reading the portion of the passage carefully* (R6), which also occurred at a very high rate (1.67); *reading a portion of the passage rapidly looking for specific information* (R7), which occurred at a moderate rate (.48); and *verifying the referent of the pronoun* as the text was read (R26), which was used at a moderate rate (.34). One other test-taking strategy that naturally groups with these *is repeating/ paraphrasing/translating the words, phrases or sentences after reading the text* (R9), which occurred at a high rate (.78). Examples that highlight strategies T5, R6, R7, R9, and R26 (as well as T22 & T28–see below):

    - "'Them.' [*Looking at passage.*] **This** 'them' **indicated that previously Paul took pictures of people exhibiting the emotions of feelings, and then he asked people what these mean; so, I think** 'them' **here should mean those pictures, because in the previous sentence pictures were mentioned.**" (C5, T1P1Q4)
    - [*Reads sentence in passage.*] "I **should read the previous sentence too.** [*Reads prior sentence.*] **I see. He asked this** 'photographs of people.' **So** 'them' **means** 'photographs.' So I choose C." (J4, T1P1Q4)
    - "So, I'll go back to the passage and read the paragraph again. [*Reads paragraph that includes pronoun.*] If you look at the sentence he says. [*Reads sentence with pronoun.*] So, he first takes some photos and asked other people about the emotions depicted in these photographs. I think the pronoun 'them' in this sentence refers to the photographs of people. Let me see what are the options in the question." (O1, T1P1Q4)
    - "Before looking at the options, I'll look at the passage first." (O6, T1P1Q4)
    - [*Reads sentence with pronoun.*] "So, 'them' would refer to the masses of ice. Let's see the options." (O7, T1P3Q8)
    - "[*Returns to passage*.] **Here the sentence starts.** [*Rereads sentence*.] **I should read the previous sentence, too.**" (J4, T1P1Q4)

- "**This question is asking what the pronoun** 'it' **stands for.** [*Locating the pronoun in the text.*] 'It' **should stand for the fossil,** 'Pakicetus.' [*Looks at options.*] **The fish, the life and the ocean are not appropriate since here the** 'it' **refers to something which bred and** 'gave birth' **on land. In the passage only** 'Pakicetus' **was mentioned this way.**" (C2, T2P3Q5)
- [*Reads question.*] **OK, let's read the related sentence in text. I am checking the grammar structure of this sentence to find what** 'them' **refers to. Analyzing where the subject and the verb ends. I think the answer is C,** valleys. **In the phrase,** 'carrying them,' **it omits its relative pronoun that function as the subject in that particular phrase.** And 'them' **refers to** valleys, **which comes right before it. I am not sure about this but I'd like to go with this.** (K4, T1P3Q8)

2. Evidence that this item type was a little challenging for subjects can be seen by the use of the following strategies: *Postponing consideration of specific options in order to consider other options first* (T16) occurred at a high rate (.59), *considering the options in turn to see how well they fit in the context* (T10) was used at a moderate rate (.38) and *rereading the question for clarification* (T1) also occurred at a moderate rate (.31). Examples that highlight strategies T16, T10, and T1:

    - [*Rereads question; reads sentence in passage.*] "OK, **here the word** 'them' **might refer to several possible things: people, emotions. Look at the options. Option** A **and** B **have people or emotions; then Options** C and D **should be eliminated.** [*Tries options in context.*] . . . 'what emotions were being depicted in them,' **since emotion and them both appear in the same clause, it is impossible for the pronoun them to refer to emotions. So, this** 'them' **should refer to people. The answer should be** B." (C8, T1P1Q4)
    - "[*Reading the sentence and predicting an answer.*] **The answer is** Glaciers. 'Carrying with them . . .' Rock? Glaciers. **How come there are no Glaciers in the choices?** 'Cold areas' **this is wrong.** 'B. masses of ice masses of ice masses of ice' 'C. valleys' 'Carrying with them . . .' **I think the** answer is **B**

‘masses of ice.’ I’ll **choose B. I’ll check D.** ‘D. rock debris’ **is wrong**.” (J5, T1P3Q8)

- [*Rereads question.*] “‘Them,’ ‘them,’ ‘them’ . . . [*Reads two sentences in text: the one including pronoun and the one prior to it.*] **In A,** ‘cold areas,’ ‘cold areas,’ ‘cold areas’? **It is not related to** ‘them.’ **In B,** ‘masses of ice’? **Not sure about this. Go to C,** ‘valleys,’ valleys **go with** ‘them,’ **the two carry some stuff and move together. So** ‘valleys’ **is not the answer. In D,** ‘rock debris’ **is not** ‘them’ **because rocks debris comes after** ‘them,’ **the pronoun. Usually, a pronoun’s referent comes before the pronoun. I think the answer is B.**” (K1, T1P3Q8)
- “OK. I’ll read the paragraph first. [*Reads sentence including pronoun and the one prior.*] So, here, it’s talking about . . . [*Rereads sentence with pronoun.*]. Here, if I look at the options, Option A is not correct because it isn’t talking about areas. Option B can be correct. Option C may be correct too. Option D is not correct because it is at the end of the sentence in the paragraph and it doesn’t refer to ‘them.’ So, it’s either Option B or C. [*Rereads sentence with pronoun.*]. So, here, I’m not sure what it refers. I’ll read the sentence again. [*Rereads sentence with pronoun.*]. So, if I put Option B instead of ‘them’ [*Rereads the sentence replacing ‘them’ with ‘rock debris.*’], it is making sense. So, I believe masses of ice carries huge quantities of rock debris. So, Option B is correct.” (O2, T1P3Q8)
- “[*Rereading the question.*] **This question is asking what the pronoun** ‘it’ **stands for.**” (C2, T2P3Q5)

3. Already reflected in the examples given above, respondents relied very heavily on their *understanding of the sentence and paragraph meaning to help them select which option was the referent for the pronoun in question* (T22), which occurred at a very high rate (1.13) *or discarding options that weren’t* (T28, which also occurred at a very high rate (1.07). Examples of strategies T22 and T28 (as well as R26):

- [*Reads question.*] [*Reads sentence in passage.*] "OK, now I remember, and I think it is a . . . in the previous sentence it mentioned 'took photographs of people.' Yeah, it means 'these photographs of people.' So, it should be 'photographs' instead of 'people.'" (C3, T1P1Q4)
- "'It' **should be the same meaning with the previous one** [*noun*]. A **should be right**." (C4, T2P3Q5)
- "OK. 'What' became suddenly public? I need to read the previous sentence to understand 'it.'" (O4, T2P2Q8).
- [*Reads question.*] "**OK, let's read the related sentence in text. I am checking the grammar structure of this sentence to find what** 'them' **refers to. Analyzing where the subject and the verb ends. I think the answer is C,** valleys. **In the phrase**, 'carrying them,' **it omits its relative pronoun that function as the subject in that particular phrase**. And 'them' **refers to** valleys, **which comes right before it. I am not sure about this but I'd like to go with this**." (K4, T1P3Q8)

Basic Comprehension-Sentence Simplification (BC-ss)

The basic comprehension-sentence simplification (BC-ss) item type 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). Examinees need to choose the option that best simplifies a targeted sentence while retaining the essential information. The accuracy rate for the 37 attempts to respond to the 6 items of this type was 30/37 = 81% (J = 70%, C = 89%, K = 67%, O = 100%). Table 13 outlines the most frequently used reading and test-taking strategies for this item type.

Table 13

Common Strategies for Item Type BC-ss

Strategy code	Strategy description	Frequency rate
T28	Discards option(s) based on vocabulary, sentence, paragraph, or passage overall meaning as well as discourse structure (depending on item type).	2.68
T5	Reads the question then reads the passage/portion to look for clues to the answer, either before or while considering options.	2.16
R6	Reads a portion of the passage carefully.	1.54
T16	Considers the options and postpones consideration of the option.	1.03
T22	Selects options through vocabulary, sentence, paragraph, or passage overall meaning (depending on item type).	.95
T12	Considers the options and selects preliminary option(s) (lack of certainty indicated).	.84
R9	Repeats, paraphrases, or translates words, phrases, or sentences—or summarizes paragraphs/passage—to aid or improve understanding.	.77
T17	Considers the options and wrestles with the option meaning.	.46
T1	Goes back to the question for clarification: Rereads the question.	.41
T19	Reconsiders or double-checks the response.	.35
T14	Considers the options and paraphrases the meaning.	.35
T24	Selects options through elimination of other option(s) as unreasonable based on paragraph/overall passage meaning.	.35
T2	Goes back to the question for clarification: Paraphrases (or confirms) the question or task.	.30

Note. Strategies are ranked and grouped according to three levels of frequency rate (# occurrences/# of items of that type): very high (≥ 1.0), high (≥ .50), and moderate (≥ .30).

In reviewing the reading and test-taking strategies that occurred at very high (≥ 1.0), high (≥ .50), and moderate (≥ .30) rates for this item type, as given in Table 13, the following notable strategy trends emerged.

1. As with the BC-pr item type, the most common strategy, occurring at a very high rate (2.16), was *going directly to the passage (before looking at the options) and reading the sentence that is being paraphrased before considering the options* (T5). Two other strategies that clustered with this strategy were *(re)reading the portion of the passage carefully* (R6) and *repeating/paraphrasing (usually through translation) the sentence in the text immediately after reading it to aid or improve understanding* (R9), which occurred at a very high rate (1.54) and high rate (.77) respectively. Examples of strategies T5, R6, and R9:

    - [*Reads question*.] "**Let's read the highlighted sentence in the passage**. [*Reads sentence in passage*.] **What does it mean by** destructive forces of nature?" (J2, (T1P3Q9)
    - [*Rereads sentence to be summarized*.] "**He did it for money, and he did not develop machines**. Exhibitors **purchase only one machine so he refused to develop a new machine**." [*Goes to options*.] (J3, T2P2Q3)
    - [*Rereads sentence to be summarized*.] "**I do not know the meaning of** 'epitome.' 'Destructive' **forces of nature . . . remaining . . . successfully resisting**." [*Goes to options.*] [J8, T1P3Q9)
    - "So the passage is talking about why he did not develop projection technology because he wanted to sell more other things before he developed the projection technology. Let's look at the answers." (K2, T2P2Q3)
    - "**The structure of backbone of** Ambulocetus **proves that it could swim as the modern whales do. It's about the backbone**." [*Looks at the options*.] (K7, T2P3Q10)
    - "So, Ambulocetus didn't have a fluke but it could swim by moving the portion of its body." (O3, T2P3Q10)
    - [*Rereads highlighted sentence*.] "**The hills and mountains are considered to be permanent but actually they are relatively short-lived in geological terms**." (C5, T1P3Q9)
    - "**Passage wants to say they are perceived as** permanent, **but they really are not**." (J2, T1P3Q9)

2. Subjects spent more time considering each of the options for the BC-ss items than with other BC item types. Two strategies could be linked in that they both dealt with the challenge that subjects had dealing with the options and deciding on a correct choice: postponing the decision to select or discard an option (T16) occurred at a very high rate (1.03) and making a preliminary (but uncertain) selection of an option (or two) (T12) occurred at a high rate (.84). Examples of these two strategies, T12 and T16:

    - [*Reads Option A.*] "It obviously seems like a good option. It talks about the same thing as described in the sentence. [*Reads Option B.*]. OK. That is possible, too. [*Reads Option C.*] I can eliminate this one because the highlighted sentence doesn't provide a solution. It simply states a problem. [*Reads Option D.*] Again, this is not mentioned in the highlighted sentence. It doesn't talk about a solution to the problem. So, I have Options A and B remaining. Let me read them again." (O4, T2P1Q10)
    - "**This option seems to be right. But wait until I read Options C and D**." (C7, T1P3Q9)
    - [*Reads Option B.*] "**This one should be the answer since it seems to be a good paraphrase, but let me continue to check the other two** [options]." (C5, T1P3Q9)
    - [*Reads Option A.*] "Not sure about this. Go to B. [*Reads Option B.*] Keep going to C." [*Reads Option C. Returns to reread sentence in passage.*] (K1, T1P3Q9)
    - "I think this is the best one so far, but let's look at the last one." (K2, T2P2Q3)
    - "**Then it is only between A and D that I have to pick**. [*Rereads Option A.*] **I think A includes all the necessary information, but I am going to check D one last time**. [*Rereads Option D.*] **Still D looks OK. It doesn't include all the information, so I am going with A. My gut feeling says A**." (K5, T2P1Q10)
    - [*Reads Option B.*] "**This could be right."** [*Reads Option C.*] (J2, T1P3Q9)
    - [*Reads Option A.*] "Option A looks like a good fit. [*Reads Option B.*] That's actually more accurate. Let's read the next one. [*Reads Option C.*] That's kind of accurate, too." (O7, T1P3Q9)

Also reflecting the challenge subjects had in dealing with the options and deciding on a correct choice were the strategies *wrestling to understand option meaning* (T17), *paraphrasing the option to aid in understanding it— usually through translation*—(T14) and *reconsidering/double-checking option choice* (T19), all three of which occurred at a moderate rate (.46, .35, and .35, respectively). Examples that highlight strategies T14, T17, and T19:

- [*Reads Option A.*] "**This doesn't make sense but let's leave it and move on to the next one.** [*Reads Option B.*] **I think it is close to the answer but I want to read the rest.** [*Reads Option C.*] **I should read this option again. Wait a second,** 'the same relationship of facial . . . culture' . . . 'same relationship,' I think this is similar, too. I am going to read the next option. [*Reads Option D.*] **I think this is not right. Because . . .** 'facial expressions' **they were not familiar with the expressions and the characters in stories. Wait a second, I am going to read the highlighted sentence again. It means that when they read the stories, they** 'responded,' they displayed familiar expressions. **I think A is not right because this does not mean an idea very different from what the highlighted sentence meant. B is not the answer either because they were not asked whether they were familiar with** basic emotions. [*Reads Option C again.***]** I think this is really closer to the answer. [*Reads Option D again.*] **I think this is not right because they were** familiar **with** . . . familiar expressions **was not mentioned at all. I think C is the answer.**" (K3, T1P1Q6)
- [*Reads Option A.*] "Yeah, it has the meaning, but it doesn't intend to explain how the problem was caused. [*Reads Option B.*] This was not mentioned at all. [*Reads Option C.*] It doesn't seem that it was mentioned. [*Reads Option D.*] No, it was not mentioned. Let me look again at A. [*Rereads question.*] I made a mistake; I thought it refers to the whole passage. [*Rereads highlighted sentence*.] Then this [Option A] should be the answer. **See, difficult to reverse, many people affected, large areas affected. Yeah, this is the answer.**" (C6, T2P1Q10)

- [*Reads Option B.*] "Is this focusing on whether it is familiar or not? I need to read [*Option B*] **again.** [*Rereads Option B.*] **This** [*Option B*] **doesn't seem to be the answer.** [*Reads Options C and D.*] **Is it Option B or D? D seems to be ambiguous . . . let's go back to B . . . Is it familiar with it or not? It asked to display** familiar facial expressions . . . **Am I wrong? . . . I don't like either of the answers. Let's read it again.** [*Rereads sentence*.] **I'll choose B. I'm not sure; actually I don't like either one, yet still B seems to be better.**" (K6, T1P1Q6)
- "The first one, I have to clarify it because I clicked the first one, but now I found the third one is a better choice. [*Rereads passage*.] No, that's wrong. OK, it did not have a fluke, so the third one is right." (C1, T2P3Q10)

3. Two strategies seemed to go together in that they characterize the challenge respondents had dealing with the BC-ss questions themselves: *rereading the question for clarification* (T1) and *paraphrasing the question—usually through translation— to help in understanding question intent* (T2), which both occurred at a moderate rate (.41 and .30 respectively). Examples of strategies T1 and T2:

    - [*Rereads and paraphrases/translates question*.] "**What does this sentence mean**." (C5, T1P3Q9)
    - "**Which of the following sentences best expresses the main meaning**?" (C7, T1P3Q9)
    - "What we are asked to do is to replace the highlighted sentence with one of the option sentences to keep the meaning unchanged." (C8, T1P1Q6)

4. When deciding which option to go with, the most common strategies dealt with *using the understanding of the sentence and paragraph meaning to either select options* (T22), which occurred at a high rate (.95), or *discard options* (T28), which occurred at a very high rate (2.68). Similarly, selecting an option by *eliminating other options based on sentence meaning* (T24) occurred at a moderate rate (.35). The following are examples of strategies T22 and T28:

    - [*Reads Option A.*] "**Variety of machine. B. Edison refused** . . . This is not so related. **C. Edison did not want** . . . This is close meaning . . . D. Edison would

not . . . **No the passage did not say it like this.** [*Chooses C.*] **Yes, that's right. He wanted to sell a lot of machine so he did not develop any more.**" (J3, T2P2Q3)

- "A. Whigs were able to attract support only . . . 'Only' **is already wrong**. . . . wealthiest party . . . **I do not think the sentence mentioned** 'wealthiest party' **although the passage said** 'languishing.' 'Only' **is wrong.** [*Reads Option B.*] 'Rural area' **and** 'urban area' **are the same so this is wrong.**" (J5, T1P2Q11)
- [*Reads Option A.*] "**The first part of the sentence where it says that they are popular among the wealthy people is right. But the word** 'only' **makes me hesitant. Usually I tend to think that sentences with these kinds of extreme words,** 'only' **or** 'must,' **are not the answer. Also, its structure, the** 'because' **part is not correct either. The** 'because' **part is not the reason that caused Whig's popularity among the wealthy. So it is not the answer. In 'B,'** [*Reads Option B again.*], **its discourse structure looks parallel, I mean not led by cause and effect like in Option A. So I'll keep this one and move to C. C is . . .** [*Reads Option C again.*] **I think this is right but . . . Well, C is not right because it shows the structure of A causes B. It is not parallel.**" (K4, T1P2Q11)
- [*Reads Option A.*] "This is incorrect because Whigs were strongest in most of the areas of the open market. [*Reads Option B.*] This is not correct because Whigs also were powerful in these areas based on the paragraph. [*Reads Option C.*] This option is too limited because the paragraph talks about other things also. [*Reads Option D.*] This is exactly the right answer because this is what the paragraph is talking about." (O2, T1P2Q11)
- [*Reads option*.] "I don't know this is true or not, but this is not the point of the last sentence." [*Goes to next option*.] (C3, T1P2Q11)
- "'Western culture' **is not mentioned anywhere, so I'll eliminate this.**" (J8, T1P1Q6)

Basic Comprehension-Factual Information (BC-f)

The basic comprehension-factual information (BC-f) item type is intended to "measure examinees' ability to identify responses to questions about 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" (ETS, 2003, p. 10). The accuracy rate for the 64 attempts to respond to the 23 items of this type was 52/64 = 81% (J = 69%, C = 94%, K = 75%, O = 88%). Table 14 outlines the most frequently used reading and test-taking strategies for this item type.

Table 14

Common Strategies for Item Type BC-f

Strategy code	Strategy description	Frequency rate
T5	Reads the question then reads the passage/portion to look for clues to the answer, either before or while considering options.	2.89
T28	Discards option(s) based on vocabulary, sentence, paragraph, or passage overall meaning as well as discourse structure (depending on item type).	1.86
R6	Reads a portion of the passage carefully.	1.83
T16	Considers the options and postpones consideration of the option.	1.45
R9	Repeats, paraphrases, or translates words, phrases, or sentences—or summarizes paragraphs/passage—to aid or improve understanding.	1.23
T22	Selects options through vocabulary, sentence, paragraph, or passage overall meaning (depending on item type).	.95
T1	Goes back to the question for clarification: Rereads the question.	.81
T2	Goes back to the question for clarification: Paraphrases (or confirms) the question or task.	.55

(Table continues)

Table 14 (continued)

Strategy code	Strategy description	Frequency rate
T12	Considers the options and selects preliminary option(s) (lack of certainty indicated).	.50
R10	Identifies an unknown word or phrase.	.45
T14	Considers the options and paraphrases the meaning.	.39
T24	Selects options through elimination of other option(s) as unreasonable based on paragraph/overall passage meaning.	.34

Note. Strategies are ranked and grouped according to three levels of frequency rate (number of occurrences / number of items of that type): very high (≥ 1.0), high (≥ .50), and moderate (≥ .30).

In reviewing the reading and test-taking strategies that occurred at very high (≥ 1.0), high (≥ .50), and moderate (≥ .30) rates for this item type, as given in Table 14, the following notable strategy trends emerge:

1. This item type required subjects to read and understand a key point in a specific paragraph. Consequently, the most common strategy for these types of items was *reading the portion of the passage indicated by the question and looking there for clues to the answer* (T5), which occurred at a very high rate (2.89). Closely related to this were the reading strategies *reading the paragraph in question carefully* (R6), occurring at a very high rate (1.83), and *summarizing or paraphrasing—often by translating into their L1—portions of the passage* (R9) to help them get a better grasp of what was being discussed, also occurring at a very high rate (1.23). Examples that highlighted strategies T5, R6, and R9 include the following:

    - [*Rereads paragraph.*] "**So, water evaporates and the salts are left behind. This is a bad thing.** [*Reads Options A–D.*] **Yes, this is the answer. The paragraph says** 'salt' **so I choose 'D.'**" (J7, T2P1Q8)
    - "I don't know what [the question] is talking about . . . 'border areas.' Let's read the passage first. [*Rereads paragraph.*] The passage is talking about how

population increases and how population can cause desertification." (K2, T2P1Q4)

- [*Rereads paragraph.*] "**What is the cause of the loss of natural vegetation for soil . . . and it seems to be different things. Look at the answers now.**" (K8, T2P1Q2)
- [*Reads paragraph.*] "The first sentence of the paragraph mentioned that the reduction of vegetation result in . . . **so naturally it has much to do with the question. In other words,** the loss of soil's ability to absorb substantial quantity of water, **which means reduced vegetation causes decrease of the soil's ability to take in water. So, let me go back to the question and choices . . .**" (C2, T2P1Q2)
- [*Reads paragraph.*] "**This paragraph talks about the impact of moving plates on the formation of mountains.**" (C7, T1P3Q2)

A related reading strategy that occurred at a moderate rate (.45) was *identifying unknown words or phrases* (R10) in the passage (or an option) as they read or reread it. Examples of strategy R10:

- "**I am wondering about** peepshows. **I have been thinking, but I don't understand this yet. Maybe it is prefilms before the films.**" (J3, T2P2Q10)
- "**I don't know what** runoff **means.**" (K8, T2P1Q2)

2. As with the other BC item types, many respondents found these types of items a little challenging. Three strategies seemed to cluster together in the respondents' efforts to deal with this item type: *postponing the decision to select or discard an option* (T16), which occurred at a very high rate (1.45), *making a preliminary (but uncertain) selection of an option (or two)* (T12), which occurred at a high rate (.50), and *paraphrasing the option to aid in understanding it—usually through translation*—(T14), which occurred at a moderate rate (.39). Examples of strategies T12, T14, and T16:

   - "**I think this** [option] **is wrong too but let's go back to the paragraph again.**" (J2, T1P1Q5)

- [*Reads Option D.*] "**So**, D's 'sand' **is wrong**. [*Reads Option A.*] A's 'glacial activity' **is the** 'result' **of cold area**. [*Reads Option B.*] **And about** B's '**rock**' . . . [*Rereads paragraph.*] 'Wind' **and** 'sand' **is mentioned here**. [*Reads Option A.*] Glacial **means 'glacial' . . . This is mentioning the holistic idea. Maybe this is the answer. B, C, and D are wrong, so A**." (J8, T1P3Q10)
- "**A is** [*Reads Option A.*]. **I don't know this clearly. B means that people become unfairly rich. C is** [*Reads Option C.*] **I don't know this one either. D is** [*Reads Option D.*] **Mmm, let's go back to the sentences in the text**." (K1, T1P2Q3)
- [*Reads Option A.*] "Not sure yet but keep going. [*Reads Option B.*] Not sure yet but keep going. [*Reads Option C.*] **This is not right because it doesn't make sense**. [*Reads Option D.*] Not sure. **Let's read the related sentence again**. [*Reads sentence from passage*.] [*Reads Option D.*] **A is not right. B and C are not either. B and C were not mentioned in the passage. And A is** . . . [*Reads Option A.*] **I think D is not the answer because** 'reduce emotional response' **is wrong; I mean it is not mentioned. So I think the answer is A**." (K3, T1P1Q11)

3. Even more than with the BC-ss item types, many respondents also made a specific effort to make sure they understood the task required of them by the BC-f item types—or at least felt the need to confirm what it was they were supposed to do. Two strategies seemed to go together with respect to the subjects' handling of the questions themselves: *rereading the question for clarification* (T1) and *paraphrasing the question—often through translation—to help in understanding question intent* (T2), both occurring at a high rate (.81 and .55 respectively). Here are examples of strategies T1 and T2:

    - [*Reads the question twice*.] "OK, I'll look at the options now. Option A doesn't look good. Option B looks good. Option C looks good as well. Option D, this may be OK. I'll look at the question again." [*Reads the question a third time*.] (O2, T1P2Q7)
    - [*Reads the question. Reads the options*.] "Option C may be possible. So, I'll check it first. They must be a cause and a result of erosion. [*Rereads the*

paragraph.] So, here this option is not correct because tree roots can't be good and bad at the same time. I need to reread the question. [*Rereads the question*.] I must read the whole paragraph to find the correct answer." (O2, T1P3Q10)

- [*Reads the question. Reads the paragraph. Reads the question a second time. Reads Options A-C. Reads the question a third time. Reads Option D*.] "I think this one is pretty good. Let me read the question again." [*Reads the question a fourth time*.] (O3, T2P1Q8)
- [*Rereads question*.] "**It's about the early exhibitors. What role did they play?**" (K7, T2P2Q7)
- [*Reads the question*.] "So, what are the similarities between these species? OK." (O3, T2P3Q4)

4. When deciding which option to go with, the most common strategies dealt with *using the understanding of the sentence and paragraph meaning to either select options* (T22), which occurred at a high rate (.89), or *discard options* (T28), which occurred at a very high rate (1.86). Examples of strategies T22 and T28:

    - "**Let's go back to the options. A looks close according to the content I just read. B looks wrong because I don't remember reading about conflict . . . but let's keep it aside. C is not related at all. D is about the party, but the paragraph did not emphasize preventing.**" [*Selects A.*] (K4, T1P2Q4)
    - "**I think the answer is B based on context, because the question is about cause and result. In A,** 'glacier activity' **cannot be both cause and result. Rather it is causing something. I think B can be both. But in C, I think it can be only cause. In D, sand can be dependent upon wind so . . . I am trying to figure why it cannot be both. I'd like to change my answer to D, sand.**" (K4, T1P3Q10)
    - "The Option A, this might be the correct choice, because the writer is talking about an experiment while people are watching cartoons. The Option B, I don't think it is the answer. The Options C and D are irrelevant. But I need to decide on Options A and B. Let me read the passage again. [*Reads paragraph.*] I think the Option A is more appropriate because this information comes right after the

sentence. [*Reads sentence.*] It looks like the author wants to provide some verification of the facial feedback hypothesis and it is related with the experiment on people while they are watching some cartoons. So, Option A is the correct answer." (O1, T1P1Q8)

- [*Reads Option A.*] "Yes, exactly. I remember this from the passage because I thought that was a very interesting idea." [*Selects Option A.*] (O7, T1P1Q7)
- "**A should be eliminated because it was mentioned in the text; also when it was mentioned in the later part of the text, it was mentioned as a destructive force. So, A shouldn't be the answer. B slowing down of volcanic activity . . . I think this is the opposite argument;** slowing down, **slowing down of volcanic activity leads to the formation of mountain; no, I think it shouldn't be right. Quite opposite. It should be frequently occurring of volcanic activity that causes the formation of mountains.**" (C7, T1P3Q5)

Similarly, selecting an option by *eliminating other options based on sentence meaning* (T24) occurred at a moderate rate (.34). Examples of strategy T24:

- "**A isn't right. I think the next paragraph explains about it so it shouldn't be right. And B isn't right because they didn't live at the same time. There were 12 million time difference. It couldn't support the heavy thing. It mentions that the** Basilosaurus **could not have walked on land. C could be right, but isn't mentioned in the paragraph. So, I'll choose D.**" (K7, T2P3Q7)
- "Option A doesn't look good . . . [*Rereads question and rereads paragraph*.] . . . So, Option B is not correct. [*Rereads Options C and D.*] . . . Option D, no the paragraph doesn't talk about that at all; so the correct answer must be Option C." (O2, T1P2Q7)

Basic Comprehension-Not/Except (BC-n/e)

This item type 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. 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" (ETS, 2003, p. 12). The accuracy rate for the 26 attempts to respond to the 3 items of this type was 22/26 = 85% (J = 100%, C = 67%, K = 71%, O = 100%). Table 15 outlines the most frequently used reading and test-taking strategies for this item type.

Table 15

Common Strategies for Item Type BC-n/e

Strategy code	Strategy description	Frequency rate
T5	Reads the question then reads the passage/portion to look for clues to the answer, either before or while considering options.	2.50
R6	Reads a portion of the passage carefully.	2.23
T28	Discards option(s) based on vocabulary, sentence, paragraph, or passage overall meaning as well as discourse structure (depending on item type).	1.92
T16	Considers the options and postpones consideration of the option.	1.35
T22	Selects options through vocabulary, sentence, paragraph, or passage overall meaning (depending on item type).	.88
T12	Considers the options and selects preliminary option(s) (lack of certainty indicated).	.81
R9	Repeats, paraphrases, or translates words, phrases, or sentences—or summarizes paragraphs/passage—to aid or improve understanding.	.58
T1	Goes back to the question for clarification: Rereads the question.	.54
T14	Considers the options and paraphrases the meaning.	.46
T24	Selects options through elimination of other option(s) as unreasonable based on paragraph/overall passage meaning.	.46
T19	Reconsiders or double-checks the response.	.31
T2	Goes back to the question for clarification: Paraphrases (or confirms) the question or task.	.31

Note. Strategies are ranked and grouped according to three levels of frequency rate (# occurrences/# of items of that type): very high (≥ 1.0), high (≥ .50), and moderate (≥ .30).

In reviewing the reading and test-taking strategies that occurred at very high (≥ 1.0), high (≥ .50), and moderate (≥ .30) rates for this item type, as given in Table 15, the following notable strategy trends emerge:

1. This item type required subjects to read and understand key points either in a specific paragraph or across the whole passage, recognizing which one of four options was not a key point. Following the same pattern as the other BC item types, the most common strategy for these types of items was *reading the portion of the passage indicated by the question and looking there for clues to the answer* (T5), which occurred at a very high rate (2.50). Closely related to this were the reading strategies *reading the paragraph in question carefully* (R6), which also occurred at a very high rate (2.23), and *summarizing or paraphrasing—often by translating into their L1—portions of the passage* (R9) to help them get a better grasp of what was being discussed, which occurred at a high rate (.58). Examples that highlight strategy T5, R6, and R9 include the following:

    - [*Reads Paragraph 6.*] "So, here, it looks like the Democrats don't have a big support from the business and commercial classes. So, I'll look for the business and commercial classes. [*Reads Option A.*] It isn't that. [*Reads Option B.*] No, it isn't that either. [*Reads Option C.*] It looks like this one is correct. OK. I'll look for the business and commercial classes. OK. I'll try to find rising entrepreneurs in the paragraph. OK. [*Rereads paragraph.*] It looks like rising entrepreneurs means the business and commercial classes. So, here the Whigs were very powerful. And the paragraph says the rising entrepreneur support the Democrats very much. And if we look at the other parts of the paragraph, it says: Whigs appealed to planters like themselves. Right. I'll change my answer because here it says Democrats, attracted rising entrepreneurs. So, in the paragraph, it talks about the planters involved in the international trade. So, Option B is correct." (O2, T1P2Q11)
    - "**This paragraph, I didn't read it to the end since it is too long; instead I just read the first sentence.** [*Rereads paragraph.*] **It is not until here that the D-party is mentioned.** [*Continues reading.*] **This word means someone who creates an enterprise . . . ok, leave it alone.** [*Continues reading.*] **Seems only**

this part of the whole paragraph talks about the D-party . . . oh, no, there's another where it is mentioned. The last sentence of the paragraph also mentions the D-party." (C7, T1P2Q10)

- [*Rereads paragraph.*] "**If he made a projector, exhibitors would purchase only the machine . . .** " (J1, T2P2Q2)

2. When respondents were deciding which option to select, the most commonly reported strategies included *using the understanding of the sentence and paragraph meaning to either select options* (T22), which occurred at a high rate (.88), or *discarding options* (T28), which occurred at a very high rate (1.92). Examples that highlight strategies T22 and T28:

    - [*Reads Options A-D.*] "**I already forgot what I read before but D is correct since it appeared where I just read.** [*Reads Options C, B, A.*] **B's** 'global warming' **was mentioned briefly although it was not so important. So B is correct, too. So, either A's** 'soil erosion' or **C's** 'insufficient irrigation' **is wrong.** [*Reads Paragraph 9.*] **Well, the paragraph that I have just read mentioned about** 'overirrigation that' **bring** 'salt' **to the surface of the ground, so C is talking about the opposite thing. So C is the answer.** [*Reads Paragraph 9.*] **So** 'overirrigation' **is a cause and C is opposite. Question is** 'except' **so C is the answer.**" (J7, T2P1Q9)
    - [*Reads Paragraph 6.*] "It says Four specific activities . . . and overirrigation. So it talks about insufficient irrigation. It also did talk about human activity and global warming. It also talked about soil erosion and how they can absorb water. But it didn't talk about the livestock . . ." (K2, T2P1Q9)
    - "**D is correct because it is mentioned. C is correct, too.**" (K3, T1P2Q10)
    - [*Reads Option C.*] "No, it doesn't say that they were popular subjects of the films. Yes, they were popular, but we can't infer that they were the most popular." (O4, T2P2Q1)
    - "Yeah, right, OK. The second group actually support the Whig Party. Although I don't quite get the meaning, like what's kind of group of people supporting

Whigs or what group of people supporting Democrats, but the text informs that the second group doesn't support the Democrats." (C3, T1P2Q10)

Similarly, selecting an option by *eliminating other options based on sentence/paragraph meaning* (T24) occurred at a moderate rate (.46). Examples focusing on strategy T24:

- [*Reads Option A.*] "This is something that the Democrats supported. So, this is right. [*Reads Option B.*] I didn't read about it there. [*Reads Option C.*] Yes, it's mentioned in the paragraph. [*Reads Option D.*] Yes, this is too in the paragraph. So, Option B is the only one that is not mentioned in the paragraph. So, I'll go with Option B as my answer to this question." (O6, T1P2Q10)
- "**A is right because the text mentions something like this, especially about farmers. B is . . . not sure I want to go back to text.** [*Rereads paragraph.*] **Here, I think D is mentioned so it is not the answer. C is also mentioned. Must be B.**" (K4, T1P2Q10)
- "**The answer should be B because it is the only one that was not mentioned in the text while A, C, and D can all be found in the text. Only B can't be found. So, the answer should be B.**" (C7, T1P2Q10)

3. As with the other BC item types, many respondents found these types of items a little challenging. Several strategies can be grouped together as reflective of the challenge this item type presented. *Postponing the decision to select or discard an option* (T16) occurred at a very high rate (1.35), *making a preliminary (but uncertain) selection of an option (or two)* (T12) occurred at a high rate (.81), and *paraphrasing the option to aid in understanding* it—*usually through translation*—(T14) occurred at a moderate rate (.46). *Reconsidering or double-checking a chosen response* (T19) also occurred at a moderate rate (.31). Examples that highlight strategies T12, T14, T16, and T19:

   - "C. Prizefights . . . **What is** prizefights? **Where did I see this word?** [*Rereads paragraph.*] **Here. But I do not think** . . . C . . . subjects for films. **Also, the passage did not say C's** 'popular.' **So, I think this is the answer.** [*Selects Option C.*] **Let's check again.**" (J3 T2P2Q1)
   - [*Reads Option B.*] "'**Planters . . .'** Planters? [*Rereads Option B.*] 'Rising . . .' **Entrepreneurs? I am not sure yet. Not clear yet. Maybe I should go back to**

the paragraph again. Planters? Planters? **Where was this word in the paragraph?** Planters? International**? I don't find anything related to** 'planters.' **Well, since I don't find B-related content in the paragraph, I am going with B. I think this is the answer.**" (K1, T1P2Q10)

- [*Reads Option A.*] "OK. This makes sense because workers unhappy with the system are the ones isolated from the current system. [*Reads Option B.*] I think that's not acceptable because it says the Democrats are supported by the farmers isolated from the markets rather than the ones involved in the international trade. So, Option B is the correct answer. But, I will look at the remaining options." (O1, T1P2Q10)
- [*Reads Option C.*] "Oh yeah, it was talking about over irrigation. So, how come it is insufficient? [*Reads Option D.*] Let's skim the passage for it. [*Rereads Paragraph 6.*] So, Option C is different than over irrigation. It's totally in a different continuum. But, how about Option D? I haven't found it in the text yet. Let's skim a little bit more. [*Reads Paragraph 7.*] Here, it says it right here. So, Option C is the correct answer for this question. It doesn't talk about insufficient irrigation but over irrigation." (O3, T2P1Q9)
- [*Reads Option A.*] "**An individual watches a movie**." (J1, T2P2Q1)
- "**I think this** [*Option C*] **is the answer; let's check again**." [*Rereads paragraph.*] (J3, T2P2Q1)
- "**According to the paragraph, I think I will choose D because it didn't mention that all films were short**. [*Rereads paragraph.*] **Oh, it did mention that it was short, so it shouldn't be the answer. Then I will choose C, because A and B, I'm sure they were mentioned in the text**." (C4, T2P2Q1)

4. Like the previous two BC item types, many respondents needed to confirm what it was they were supposed to do. Two options can be grouped together reflecting the challenge subjects had with dealing with the <u>questions</u>, including *rereading the question for clarification* (T1), which occurred at a high rate (.54), and *paraphrasing the question—usually through translation—to help in understanding question intent* (T2), which occurred at a moderate rate (.31). Examples of strategy T1 and T2:

- [*Reads question*.] "Kinetoscope? Kinetoscope?" [*Rereads question*.] (J1, T2P2Q1)
- [*Reads question*.] "**So I should find something wrong.**" (J3, T2P2Q1)
- [*Reads question*.] "**So I should find something that does not** contribute to . . ." (J6, T2P1Q9)
- [*Reads the question twice*.] "**This question says that all the following are mentioned but one.**" (C2, T2P1Q9)

Inferencing Items: Basic Inference (I)

The basic inference (I) item type 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). Examinees are asked to identify which of four options constitutes an appropriate inference based on explicit information in the text. The accuracy rate for the 45 attempts to respond to the 6 items of this type was 25/45 = 56% (J = 78%, C = 73%, K = 31%, O = 50%). Table 16 outlines the most frequently used reading and test-taking strategies for this item type.

Table 16

Common Strategies for Item Type I

Strategy code	Strategy description	Frequency rate
T5	Reads the question then reads the passage/portion to look for clues to the answer, either before or while considering options.	2.38
T28	Discards option(s) based on vocabulary, sentence, paragraph, or passage overall meaning as well as discourse structure (depending on item type).	2.24
R6	Reads a portion of the passage carefully.	1.82
T16	Considers the option(s) and postpones consideration of the option(s).	1.13
R9	Repeats, paraphrases, or translates words, phrases, or sentences—or summarizes paragraphs/passage—to aid or improve understanding.	.93

(Table continues)

Table 16 (continued)

Strategy code	Strategy description	Frequency rate
T22	Selects option(s) through vocabulary, sentence, paragraph, or passage overall meaning (depending on item type).	.93
T12	Considers the option(s) and selects preliminary option(s) (lack of certainty indicated).	.73
T1	Goes back to the question for clarification: Rereads the question.	.69
T14	Considers the option(s) and paraphrases the meaning.	.58
T2	Goes back to the question for clarification: Paraphrases (or confirms) the question or task.	.51
T3	Goes back to the question for clarification: Wrestles with the question intent.	.42
T24	Selects option(s) through elimination of other option(s) as unreasonable based on paragraph/overall passage meaning.	.38

Note. Strategies are ranked and grouped according to three levels of frequency rate (number of occurrences / number of items of that type): very high (≥ 1.0), high (≥ .50), and moderate (≥ .30).

In reviewing the reading and test-taking strategies that occurred at very high (≥ 1.0), high (≥ .50), and moderate (≥ .30) rates for this item type, as given in Table 16, the following notable strategy trends emerge.

1. Three of the four top strategies used in this item type reflected subjects' efforts to understand and use the ideas in the passage to choose the correct option from the test item. *Returning to the passage to look for clues to the answer* (T5) was the most common strategy choice, used at a very high rate (2.38). *Discarding and selecting options based on paragraph/overall passage meaning* (T28; T22) occurred at very high (2.24) and high (.93) rates, respectively. Examples of strategies T5, T22, and T28:

    - "I know that the question here is referring to something within the Whig Party. So, I'm gonna go back to the passage and read more." (O6, T1P2Q9)
    - "OK. So, I'll go back to the passage and read it." (O5, T2P3Q2)

- [*Reads question.*] "Here we have to find sea otters. [*Returns to paragraph in passage.*] Here are some examples." (C1, T2P3Q2)
- "**So, the height should be an indication of its age; in this case, the first option is not right. The second shouldn't be the answer because this mountain** [Caledonian] **has been shortened.**" (C5, T1P3Q3)
- "**Yes, the third option is right because the text did mention that they were the same size 400 million years ago.**" (C5, T1P3Q3)
- "**This one** [*A*] was **not mentioned though it is true. The author didn't mention it.**" (C6, T2P1Q11).
- [*Reads Option B.*] "**This one should be immediately eliminated because it is completely irrelevant; also it has not been mentioned.**" (C7, T1P2Q9)
- [*Reads Option B.*] "No. It doesn't talk about numbers. It's wrong, too. [*Reads Option C.*] The passage doesn't say that they only lived in the sea. So, it's wrong." (O5, T2P3Q2)
- [*Rereads Option D.*] "I think this one is correct because they didn't leave many fossil remains; we can't envision what they looked like." (O3, T2P3Q2)
- [*Reads Option A.*] "This option says it's difficult to imagine, and the paragraph says it's not easy to envision. They're opposite things. So Option A is wrong." (O5, T2P3Q2)
- "**D is regional interest. I want to go back to the text.** [*Rereads paragraph.*] **Wait, on the last part of Paragraph 5, it mentions about their conflict regarding regional differences. Well, then, A is more for general and D is more specific difference within the party, so this is the answer.**" (K4, T1P2Q9)

A less common strategy, but one that also highlights the subjects' efforts to select the correct option based on their understanding of the text is the *selection of an option through the elimination of other options as unreasonable based on passage meaning* (T24), which occurred at a moderate rate (.38). Examples of strategy T24:

- [*Rereads Option D.*] [*Rereads Option C.*] "I think it's irrelevant. [*Rereads Option B.*] There is no information about that. [*Rereads Option A.*] It's not

related to public liberty. It's related to moral and religious issues. So, Option D is the only one addressing religious issues." [*Selects Option D.*] (O1, T1P2Q9)

- [*After rejecting Options A-C, reads Option D.*] "This makes sense, and it's the only option remaining, so my answer is Option D." (O5, T2P3Q2)
- "**So, A is wrong; it did not mention C; B is not right either . . . so it should be D.**" (C2, T2P3Q8)
- "**C is obviously not right because it says they lived on land and at sea. A is not right either because it says it's easy to imagine. D could be right, because it's about fossil . . . Oh! It's clearly not right because D says it did not leave . . . So, it's B.**" (K7, T2P3Q2)
- "Option A is not related to what the author is saying. Again, he doesn't talk about change either, so Option B is wrong. Option D is also wrong because the author doesn't say such a thing. So, I have Option C as the correct answer." (O8, T2P1Q11)

2. Other strategies reflected the need for subjects to have a good grasp of the overall meaning of the paragraph addressed by the question. Subjects *returned to the passage to (re)read carefully* (R6) and *summarized or paraphrased*—usually through translation—(R9) the paragraph(s) in the passage referred to by the question at very high (1.82) and high (.93) rates respectively. Examples of strategies R6 and R9:

    - [*Rereads paragraph.*] "**From the second paragraph, it can be inferred that the higher the mountain is the younger it is.**" (C5, T1P3Q3)
    - [*Rereading paragraph in passage*.] "**This explains reasons.**" (C6, T2P1Q11)
    - [*Rereading paragraph in passage*.] "**Oh, here, the political beliefs should refer to how this party viewed the political issues differently; for example, in the areas of economy, industry or agriculture, what views did they hold? Yes, it should be interpreted this way.**" (C7, T1P2Q9)
    - [*Rereading paragraph in passage*.] "**So, it's a good thing that they have strong legs.**" (O5, T2P3Q9)
    - [*Rereading paragraph in passage*.] "Whigs **in the northern sections believed the government to support the moral welfare of the country.**" (K1, T1P2Q9)

3. This item type also proved to be among the most difficult of all the item types for subjects to understand. *Rereading the question for clarification* (T1) and *paraphrasing the question* (T2) occurred at a high rate (.69 and .51 respectively), and *wrestling with the question intent* (T3) occurred at a moderate rate (.42), which is the highest rate for all the item types. Examples of strategies T1, T2, and T3:

    - [*Reads the question. Rereads the question*.] "OK, I'll read the question again before looking at the paragraph. [*Reads the question a third time*.] So, I'll look for differences in the strength of political beliefs of groups within the Whig party. [*Rereads Paragraph 5 in passage. Rereads the question a fourth time. Reads Options A and B.*] I'll read the question one more time." (O2, T1P2Q9)
    - [*Rereads the question*.] "**According the second paragraph, what should the modern Himalayas be?**" (C5, T1P3Q3)
    - [*Rereads the question*.] ". . . inferred **means concluded from, concluded about what, concluded about political beliefs, the strength of political beliefs, oh, it's about the W-party. The W-party has different political belief groups?**" (C7, T1P2Q9)
    - "**I guess, my understanding of the question is wrong. None of the options talks about positive things. I'll reread it.**" (O5, T2P3Q9)
    - [*Reads question*.] "OK. I need to read the question one more time. [*Rereads question*.] I still don't understand what the question is asking. [*Rereads question again.*] I still don't understand." (O6, T1P2Q4)
    - "**This question is troublesome because the sentence is too long to get the main idea of it; it has too many clauses.** [*Rereads question.*] [*Returns to text.*] [*Rereads question.*] **This is my fifth time to read this question itself, but I'm still confused by it. And I have no clue what it is asking.**" (C8, T1P2Q9)
    - "**This question seems to be a little confusing.**" (C7, T1P2Q9)
    - "**But, I really think the question is confusing, and I don't really understand what is asking; the way it is expressed really confuses me.**" (C7, T1P2Q9)
    - "OK, I'm having difficulty relating the question to the options." (O2, T1P2Q9)

- [*Rereads the question four times*.] "**I do not understand the question.**" (J4, T1P2Q9)

4. Also reflecting the challenge of this item type, subjects *postponed consideration of option(s)* (T16) at a very high rate (1.13) before making final decisions. Similarly, subjects used the strategies of *paraphrasing option meaning* (T14), to confirm they had a good grasp of the option, and *making a preliminary choice* (T12), showing the tentativeness of their decision, at high rates (.58 and .73, respectively). Examples of strategies T12, T14, and T16:

    - "It looks like this one [*Option C*] because it says [*going to the text*] it has been estimated an additional one fourth, is threatened. Let me see the following part of the passage . . . [*Rereads and paraphrases sentence in text.*] Who knows? [*Looks at Options A and D*.] In all areas [*Option D*], it might be, but it didn't say that." (C6, T2P1Q11)
    - [*Reads Option C.*]. "**I'm not sure about this one.** [*Reads Option D.*] **I don't know this one either.**" (O5, T2P3Q9)
    - [*Reads Option B*.] "**It will change in the future. The factors influencing desertification will change in the future. [***Moves on and reads Option C.*] **Desertification will continue to increase.** [*Moves on and reads Option D*.] **Desertification will soon occur in all areas of the world.**" [*Returns to and rereads passage*.] (J1, T2P1Q11)
    - [*Reads Option D.*] "The paragraph isn't talking about the regional interests. [*Rereads Option A.*] Are they really talking about public liberty? . . . so here it's not about the differences very much." (O2, T1P2Q9)
    - [*Reads Option A*.] "**Was this mentioned? I'll come back to this."** [*Reads Option B*.] (J2, T1P3Q3)
    - [*Reads Options A–D. Rereads question. Rereads Options A–D*.] ". . . **I'll choose Option D, although I do not have much confidence. I'll look again.**" (J4, T1P2Q9)
    - [*Reads Option A*.] "**I don't know. Let's leave this for now.**" (J5, T1P2Q9)

- [*Reads Option D.*] "**It should be** regional interests**.** [*Rereads Paragraph 5. Rereads Option D.*] **Yes,** it reflects the northern section." (J5, T1P2Q9)
- [*Reads Option B.*] "**I don't think the passage talks about** change. **I'll go on.** [*Reads Option C.*] **I'll come back to this.** [*Reads Option D.*] **All area is too much to say.**" (J6, T2P1Q11)

Inferencing-Rhetorical Purpose (I-rp)

The Inferencing-rhetorical purpose (I-rp) item type 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). The accuracy rate for the 27 attempts to respond to the 3 items of this type was 23/27 = 85% (J = 57%, C = 100%, K = 100%, O = 86%). Table 17 outlines the most frequently used reading and test-taking strategies for this item type.

Table 17

Common Strategies for Item Type I-rp

Strategy code	Strategy description	Frequency rate
T5	Reads the question then reads the passage/portion to look for clues to the answer, either before or while considering options.	2.27
T16	Considers the option(s) and postpones consideration of the option(s).	2.19
T28	Discards option(s) based on vocabulary, sentence, paragraph, or passage overall meaning as well as discourse structure (depending on item type).	2.08
R6	Reads a portion of the passage carefully.	1.66
T22	Selects option(s) through vocabulary, sentence, paragraph, or passage overall meaning (depending on item type).	.96
R9	Repeats, paraphrases, or translates words, phrases, or sentences—or summarizes paragraphs/passage—to aid or improve understanding.	.85

(Table continues)

Table 17 (continued)

Strategy code	Strategy description	Frequency rate
T12	Considers the option(s) and selects preliminary option(s) (lack of certainty indicated).	.58
T1	Goes back to the question for clarification: Rereads the question.	.58
T14	Considers the option(s) and paraphrases the meaning.	.46
T17	Considers the option(s) and wrestles with option(s) meaning.	.38
T6	Predicts or produces own answer after reading the portion of the text referred to by the question.	.31
T2	Goes back to the question for clarification: Paraphrases (or confirms) the question or task.	.31
R7	Reads a portion of the passage rapidly looking for specific information	.30

Note. Strategies are ranked and grouped according to three levels of frequency rate (number of occurrences / number of items of that type): very high (≥ 1.0), high (≥ .50), and moderate (≥ .30).

In reviewing the reading and test-taking strategies that occurred at very high (≥ 1.0), high (≥ .50), and moderate (≥ .30) rates for this item type, as given in Table 17, the following notable strategy trends emerged:

1. Because of the nature of this item type, all subjects returned to whatever part of the passage was necessary *in order to find clues to the answer* (T5), with this strategy occurring at a very high rate (2.27). Three related reading strategies are *rereading the portion of the passage* carefully (R6: 1.66), *paraphrasing—often by translation— the main idea of the sentence/paragraph* (R9: .85), and *(re)reading a portion of the passage rapidly looking for specific information* (R7: .30). Examples of strategies T5, R6, R7, and R9:

    - [*Reads question.*] "**Let me read the second paragraph.**" (K7, T2P2Q2)
    - [*Rereads passage.*] "**Weather, one of its factors is carbon dioxide.**" (C7, T1P3Q6)

- [*Reads question.*] "I'll go back to the 2nd paragraph. [*Rereads paragraph.*] So, it's talking about hearing different recorded speeches or pieces of music." (O5, T2P2Q1)
- "OK. So, they want to find out why the highlighted phrase is used in the passage. So, I'm gonna go back to the passage and try to find out." (O6, T1P1Q2)
- [*Reads Option D.*]. "OK. I'm not sure about this one, so I'm going to read more from the passage to better understand it [*Returns to passage.*]." (O6, T1P1Q2)
- [*Reads paragraph.*] "I think what the author wants to do here is to just give us another example as to how facial expressions are universally understood." (O6, T1P1Q2)
- [*Reads question.*] "**I should focus on** Phonograph parlors **and read** paragraph 2." [*Rereads paragraph.*] (J6, T2P2Q2)
- [*Reads question.*] "**Let's read the passage.**" (J1, Y2P2Q2)
- [*Rereads paragraph.*] "**If he made projectors, exhibitors would buy only one machine.**" (J1, T2P2Q2)
- "**The second paragraph tells about why Edison refused to improve the projector; that is, because in that case he could sell more machines. So, he didn't improve or make more than the one projector to sell more machines.**" (C4, T2P2Q2)
- [*Reads question; rereads highlighted portion of passage*.] "Looking back here, basically carbon dioxide is mentioned because it changes the rocks chemically and makes them weaker." (O7, T1P3Q6)

2. As with item-type "I" above, *using understanding of the paragraph meaning*, and to some extent overall passage meaning, to *discard* (T28) or *select* (T22) *specific options* were two strategies that occurred at very high (2.08) and high (.96) rates, respectively. Examples of strategies T22, and T28:

    - "**This is not the answer, either . . . why? Because . . . the passage was not about the Darwin's theory . . .**" (K6, T1P1Q2)

- [*Reads paragraph.*] "**It first mentions about the phonograph and then says something about the kinotoscope, so I guess it's about finding the similarities of those two. So A is not right.**" (K7, T2P2Q2)
- [*Reads Option B.*] "**No, it is not mentioned. It says it compares** carbon dioxide **with** carbonic acid. **No, it's not right. Actually, it says that** carbon dioxide **forms the** carbonic acid. **In other words, it says** carbon dioxide **is one of the causes of** carbonic acid, **rather than comparing them. They have a relationship of cause and effect rather than of comparison. So, B shouldn't be the answer.**" (C7, T1P3Q6)
- [*Reads Option D.*] "Well, it was improved but, is this the purpose he discussed this? I don't think so, because they were using, yeah [*returnings to passage*] . . . these Kinetoscope arcades were modeled on phonograph parlors, which means they were the same model. So, he didn't mean to illustrate this [that they were more technologically advanced]." (C6, T2P2Q2)
- [*Reads Option C. Rereads paragraph.*] "**They are comparing but not talking about which one is popular. Option C, the word** popular **is not mentioned.**" (J3, T2P2Q2)
- "**I think the paragraph explains the role of carbon dioxide, and Option A explains it exactly.**" (K4, T1P3Q6)

A less frequent but notable strategy that also reflected subjects' focus on text meaning was the *prediction of the answer after having read the portion of the text referred to by the question*—but before looking at the options (T6), which occurred at a moderate rate (.31). This was the highest rate for this strategy among all the item types. Examples of strategy T6:

- [*Reads question; rereads highlighted portion of passage.*] "**Because previously he mentioned that the expression of emotions was universal and that smiling means friendliness and approval, then he continued to show how to express unfriendliness. Let me have a look at the options.**" (C5, T1P1Q2)
- [*Reads question; rereads highlighted portion of passage*.] "**The author mentions this, obviously, in order to give an example showing that a facial**

expression can express the same emotion in different cultures. [*Begins reading options.*] **The first option . . .**" (C8, T1P1Q2)

- "**When I read these options, I was looking for the reasons as I just mentioned** [*when paraphrasing Paragraph 2.*], **but . . .**" (C4, T2P2Q2)
- [*Reads question; rereads highlighted portion of passage.*] "So, it talks about how it was modeled after phonograph parlor. Let's look at the answers." (K2, T2P2Q2)
- [*Reads question; rereads highlighted portion of passage.*] "So, it mentioned carbon dioxide because it's one of the main reasons that penetrates rocks and results in the breaking of rocks. So, let me read the options now." (O2, T1P3Q6)

3. Similar to the "I" item type discussed above, the challenge of this item type was reflected by the very high rate (2.19) of the strategy of *postponing consideration of an option* (T16) before making final decisions. *Wrestling with option meaning* (T17) also occurred at a moderate rate (.38), reflecting some of the challenge of the option choices as well. Likewise, subjects used the strategies of *making a preliminary choice* (T12), showing the tentativeness of their decision, and *paraphrasing option meaning* (T14), to confirm they had a good grasp of the option, at high (.58) and moderate (.46) rates, respectively. Examples of strategies T12, T14, T16, and T17:

    - "**This seems to be the right one . . . Let's see the last one.** [*Reads Option D.*] **This might be the answer, too. But . . . let's go back.**" (K6, T1P1Q2)
    - "I don't quite understand what this first option means. I'll go on to the next option." (C5, T1P1Q2)
    - [*Reads Option A.*] "I'm going to read it again. I didn't understand it. [*Rereads Option A.*] I'm still not sure what it means. So, I'm going to move on to the next option."
    - [*Reads Option B.*] "**I don't really understand this.**" (J1, T2P2Q2)
    - "**This option looks like the right one, but I'd better read through the rest of options.**" (C7, T1P3Q6)

- "**The third option says it provides another example of facial expression the meaning of which is widely understood. I think this might be right.**" [*Reads Option D.*] (C5, T1P1Q2)
- [*Reads Option A.*] "**Differentiate a possible meaning of a certain facial expression from other possible meanings.**" (C8, T1P1Q2)
- [*Reads Option A.*] "It might be true. [*Reads Option B.*]. I guess this is true." (O5, T2P2Q2)
- [*Reads Option B.*] "**Yes, this is the answer**. [*Reads Option D.*] **This seems OK too**. [*Rereads paragraph.*] **I now feel like A is the answer**. [*Rereads paragraph.*] [*Rereads Option B.*] **I think B is still the correct answer**." [*Selects Option B.*] (J6, T2P2Q2)
- [*Reads Options A-D.*] "I'm going back to the passage because I'm confused with the options." (O7, T1P1Q2)
- [*Reads Option C.*] "I think that looks correct to me because he is giving examples with the most known emotions expressed by people, which is the case of happiness, and the other one is the case of anger. Option C looks correct to me, but I'll read the last option to see and make sure about my choice. [*Reads Option D.*] This one also looks not bad. So, let me go back to the passage and check on the 3rd and 4th options." (O1, T1P1Q2)

4. Also like the "I" item type discussed above, this item type proved to be a challenging item type for some examinees, or at least was one that they approached with caution. *Rereading the question for clarification* (T1) occurred at a high rate (.58), and *paraphrasing the question* (T2)—often by translation—occurred at a moderate rate (.31). Examples of strategies T1 and T2:

    - [*Rereads question.*] "**It's about the phonograph.**" (K7, T2P2Q2)
    - [*Rereads question.*] "**What is the author's purpose when he mentioned showing teeth in an unhappy way**?" (C5, T1P1Q2)
    - [*Rereads question.*] "**Why did he discuss phonograph parlors**?" (C6, T2P2Q2)

- [*Reads question.*] "Phonograph . . . phono- means sound; parlor maybe means somebody, some people, or kind of something; I'll figure it out." [*Rereads passage.*] (C1, T2P2Q2)
- [*Rereads question.*] "Carbon dioxide… **is this CO2?**" (K4, T1P3Q6)

Inferencing-Insert Text (I-it)

The Inferencing-insert text (I-it) item type is intended to "measure 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 that is displayed to them" (ETS, 2003, p. 31). The accuracy rate for the 55 attempts to respond to the 6 items of this type was 50/55 = 91% (J = 94%, C = 100%, K = 92%, O = 75%). Table 18 outlines the most frequently used reading and test-taking strategies for this item type.

Table 18

Common Strategies for Item Type I-it

Strategy code	Strategy description	Frequency rate
T5	Reads the question then reads the passage/portion to look for clues to the answer, either before or while considering option(s).	1.60
R6	Reads a portion of the passage carefully.	1.51
T22	Selects option(s) through vocabulary, sentence, paragraph, or passage overall meaning (depending on item type).	.96
T28	Discards option(s) based on vocabulary, sentence, paragraph, or passage overall meaning as well as discourse structure (depending on item type).	.84
R7	Reads a portion of the passage rapidly looking for specific information	.67
R9	Repeats, paraphrases, or translates words, phrases, or sentences—or summarizes paragraphs/passage—to aid or improve understanding.	.51

(Table continues)

Table 18 (continued)

Strategy code	Strategy description	Frequency rate
T8	Predicts or produces own answer after reading questions that require text insertion (I-it types).	.49
T15	Considers the option(s) and drags and considers the new sentence in context (I-it).	.42
T1	Goes back to the question for clarification: Rereads the question.	.36
T2	Goes back to the question for clarification: Paraphrases (or confirms) the question or task.	.35
T26	Selects option(s) through discourse structure.	.31

Note. Strategies are ranked and grouped according to three levels of frequency rate (number of occurrences / number of items of that type): very high (≥ 1.0), high (≥ .50), and moderate (≥ .30).

In reviewing the reading and test-taking strategies that occurred at very high (≥ 1.0), high (≥ .50), and moderate (≥ .30) rates for this item type, as given in Table 18, the following notable strategy trends emerged:

1. Because of the nature of this item type, all subjects returned to the passage to read the indicated paragraph *in order to find clues to the answer* (T5), which would help them determine where to best insert the new sentence. This strategy occurred at a very high rate (1.60). Also reflecting the nature of the item type are the very high (1.51) and high (,67) rates of occurrence for the two reading strategies *reading a portion of the passage carefully* (R6) and *reading a portion of the passage rapidly looking for specific information* (R7), respectively. Another strategy that could be grouped with these is *predicting/producing an answer for where the new sentence should be inserted* (T8) before looking at the possible options, which occurred at a moderate rate (.46). Examples of strategies T5, T8, R6, and R7 include:

    - "**OK, let's go back to the passage with the squares.**" (K8, T2P1Q12)

- "First, I'm going to read the sentence. [*Reads the new sentence.*] Now, I'm going to read the whole paragraph to see where the given sentence may fit." (O6, T1P1Q12)
- [*Reads the question.*] "So, it looks like after this sentence there will be an explanation of those methods which are mentioned… I'll go to the passage and see where the author explains these simple methods so the sentence should come before the explanation of these methods." (O1, T1P1Q12)
- [*Reads the question and new sentence.*] "So, this sentence should be put in a place where after a section of the passage where they introduce the new party. So, I'll look for something related to the new party." (O1, T1P2Q12)
- [*Reads new sentence.*] "This economic reliance tells me immediately that it's not going to be at the beginning of the paragraph. So, they must have talked something related to it before they can say this sentence. Let me read the paragraph." (O4, T2P1Q12)
- [*Reads the new sentence.*] "There must be a problem prior to the given sentence. Let me read the paragraph." (O8, T2P3Q12)
- [*Reads the new sentence.*] "'This new party…,' **so the sentence before this one is the key.**" (J5, T1P2Q11)

As with the previous BC and I item types, another related strategy that occurred at a high rate was *paraphrasing—often by translation—the main idea of the sentence* (R9: .51). Examples of strategy R9:

- [*Rereads sentence in paragraph.*] "This is like a transition from the first stated theory or conclusion about the universality in the recognition of emotion, then it is like use some experiment or method to prove it." (C3, T1P1Q12)
- [*Rereads two sentences in paragraph.*] "**This sentence describes the universality of the recognition of emotion …The next sentence indicates that Paul took photographs, and so on**." (C5, T1P1Q12)
- [*Rereads two sentences in paragraph.*] "**So far only D-party has been mentioned. This word** Whig party **is the first time to appear here.**" (C7, T1P2Q12)

- "**Let me read the first sentence of this paragraph.** [*Rereads sentence.*] **This means in the very cold weather, how erosion takes place.**" (C7, T1P3Q11)
- [*Rereads sentence in paragraph.*] "**So that means widespread use of projection…began to hurt.**" (J1, T2P2Q12)
- [*Rereads sentence in paragraph.*] "**This is one type of** destructive force." (J2, T1P3Q11).

2. As with the other two Inferencing item-types already discussed, *using understanding of the paragraph meaning to discard* (T28) or *select* (T22) *specific options* were two strategies that occurred fairly frequently, both occurring at high rates (.84 and .96, respectively). Selecting an option specifically based on the *discourse structure* (T26) occurred at a moderate rate (.31), but was more common in this item type than in any other item type. Examples of strategies T22 and T28:

    - [*Inserts new sentence into paragraph.*] "The rest of the paragraph talks about other things, but none of them are related to economy, so this must be the correct choice." (O2, T1P2Q12)
    - [*Reads new sentence. Rereads paragraph.*] "I think the given sentence will fit perfectly at the end of the paragraph. Let me put it there and see if it makes sense. [*Inserts new sentence at end of paragraph.*] Actually, no. It will probably make much more sense if we put it after the 2nd sentence of the paragraph, the 3rd square. [*Moves new sentence to new location.*] Because the 2nd sentence talks about how other people developed new devices. So, Edison's concern for his Kinetoscope business is related to the others developing perfected projection devices." (O4, T2P2Q12)
    - "**This is** 'moreover.' **The first sentence and second sentence are connected by** 'moreover' **so I cannot insert it here.**" (J2, T1P1Q12)
    - [*Reads new sentence.*] "**So,** 'another type' **must come after** 'one type' . . . [*Reads paragraph.*] **This is one type of** 'destructive force' **so let's insert it after this sentence.**" (J2, T1P3Q11)
    - [*Reads the question.*] "**I see. So I look for when this** 'widespread use of projection technology' **is mentioned.** [*Rereads paragraph.*] **The projection is**

mentioned here. The first and second square did not talk about projection. I should insert it in the last square." (J3, T2P2Q12)

- "**So the sentence should fit here after the first sentence very well, because this first sentence already mentions the economy. The word** 'this' **means the sentence cannot be placed before the first sentence because it must have been mentioned before. Also, the second squared sentence says the consequences. All this justifies the location between the first and second sentences. This way, the first sentence says the raising of livestock is a major part of the economy, and the selected sentence describes how this economy relies on overgrazing, while the original second sentence talks about the consequences of this reliance. So it's perfect to put it here.**" (C2, T2P1Q12)
- "**It mustn't be put in front of the first sentence because it hasn't been invented yet…nor can it be put in front of the second sentence for the same reason; if it was not yet invented, how could it talk about it.**" (C6, T2P2Q12)
- "**At the second square it is already mentioning details, rather than general explanation. I don't think it will be connected smoothly here.**" (K3, T1P2Q12)

3. Reflecting the fact that many examinees needed to spend time working through the options to try and figure out the right answer, they *returned to the question to both reread* (T1) and *paraphrase—often in their L1* (T2) at moderate rates (.36 and .35, respectively). Examples of strategies T1, and T2:

    - [*Rereads the new sentence.*] "'This party' is referring to the Whigs." (O6, T1P2Q12)
    - [*Rereads the question.*] "It's about adding a sentence into the paragraph." (O8, T2P1Q12)
    - [*Rereads the question.*] "**Where should I insert this sentence?**" (J3, T2P2Q12)
    - [*Rereads the question.*] "**This question is about where I should insert the new sentence. Let's read it**." (J5, T1P2Q11)

- [*Rereads the question.*] "**Where am I supposed to insert the sentence?**" (J5, T1P3Q11)

Similarly, many examinees were careful in trying to decide on the correct option, *dragging the new sentence to different contexts to test it* (T15). Examples of strategy T15:

- [*Rereads first two sentences in second paragraph. Inserts new sentence after this and reads aloud. Moves on and reads third sentence in paragraph. Inserts new sentence after this sentence and reads aloud.*] "No. [*Reinserts new sentence between second and third sentence in paragraph and reads all three sentences aloud.*] I think this looks fine." (O1, T1P1Q12)
- [*Rereads first sentence in paragraph.*] "Let's try to put the sentence here at the end. [*Reads new sentence, then reads next sentence in paragraph.*] OK, let's put the sentence here at the end of this sentence and see how it fits." [*Moves new sentence to new location and tests.*] (O2, T1P2Q12)
- [*Inserts new sentence in first space.*] "No, it didn't fit in the 1st square. [*Inserts new sentence in second space.*] No, I don't think it fits here as well, because it doesn't look like a smooth transition. [*Inserts new sentence in third space.*] No."
- "**So, the sentence says 'new,' so it should be placed before this sentence…I should read the sentence in context to see if it makes sense…Yeah, this sentence makes a lot of sense when put here.**" (C7, T1P2Q12)
- [*Inserts new sentence in second space.*] "**If I put it in the second space, it doesn't seem to be related to the** 'simple method.' **Let's try the third one.** [*Inserts new sentence in third space. Rereads sentences in context.*] **No, D is not the answer. Look at B again.**" [*Reinserts new sentence in second space.*] (K6, T1P1Q12)

4. It's interesting to note that, unlike all the other item types, respondents didn't seem to struggle with the options as much, as *postponing consideration of an option* (T16) before making final decisions, and *making a preliminary choice* (T12), both of which occurred at a high or very high rate for most of the other item types, only occurred at a low rate for item type I-it.

Reading to Learn Items: Reading to Learn-Prose Summary (R2L-ps)

The Reading to Learn-prose summary (R2L-ps) item type 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). The R2L-ps items were meant to call on respondents to *read through the entire text* in order to select those three statements that served to describe the text in a summary fashion. The accuracy rate for the 54 attempts to respond to the 5 items of this type (three correct choices possible for each item) was 142/162 = 88% (J = 87%, C = 88%, K = 82%, 93%). Table 19 outlines the most frequently used reading and test-taking strategies for this item type.

Table 19

Common Strategies for Item Type R2L-ps

Strategy code	Strategy description	Frequency rate
T22	Selects option(s) through vocabulary, sentence, paragraph, or passage overall meaning (depending on item type).	3.30
T16	Considers the option(s) and postpones consideration of the option(s).	2.93
T28	Discards option(s) based on vocabulary, sentence, paragraph, or passage overall meaning as well as discourse structure (depending on item type).	2.67
T4	Reads the question and considers the option(s) before going back to the passage/portion.	.85
R6	Reads a portion of the passage carefully.	.70
R7	Reads a portion of the passage rapidly looking for specific information	.67
T14	Considers the option(s) and paraphrases the meaning.	.57
T19	Reconsiders or double-checks the response.	.52

(Table continues)

Table 19 (continued)

Strategy code	Strategy description	Frequency rate
T17	Considers the option(s) and wrestles with option(s) meaning.	.37
T1	Goes back to the question for clarification: Rereads the question.	.35
R9	Repeats, paraphrases, or translates words, phrases, or sentences—or summarizes paragraphs/passage—to aid or improve understanding.	.33
T2	Goes back to the question for clarification: Paraphrases (or confirms) the question or task.	.33

Note. Strategies are ranked and grouped according to three levels of frequency rate (number of occurrences / number of items of that type): very high (≥ 1.0), high (≥ .50), and moderate (≥ .30).

In reviewing the reading and test-taking strategies that occurred at very high (≥ 1.0), high (≥ .50), and moderate (≥ .30) rates for this item type, as given in Table 19, the following notable strategy trends emerged:

1. Occurring at a high rate (.85) was the strategy of *reading the option(s) before going back to the passage* (T4). This strategy was particularly common because examinees apparently felt they had a good handle on the main ideas in the passage from working through the previous items on the test. This runs counter to how examinees approached the BC and I item-types, for which the more common strategy was to return to the passage first before considering the options. Two other test-taking strategies can be clustered with this one since all three focus on how subjects dealt with the question itself: *rereading the question for clarification* (T1) and *paraphrasing (or confirming) the question or task* (T2), both occurring at moderate rates (.35 and .33, respectively). In addition, two reading strategies, both occurring at high rates, can be naturally grouped with these, *(re)reading the portion of the passage* carefully (R6: .70) and *(re)reading a portion of the passage rapidly looking for specific information* (R7: .65). Examples of strategies T4, T1, T2, R6, and R7 include:

    - [*Reads question and introductory sentence. Reads first summary sentence option.*] "**This is correct, so I choose this.** [*Reads through Options 2-6,*

selecting Option 4. After reading Option 6:] **Let's go back to the passage.**" (J1, T2P1Q13)

- [*Reads and rereads question and introductory sentence.*] "So I need to choose three. [*Reads first summary sentence option.*] I don't think this is right because it was not mentioned. [*Continues reading Options 2-6.*] I think the last should be right, too, but I think I need to read the text." (C1, T2P2Q13)
- [*Reads Option 1.*] "Duchenne, Duchenne? **Was this name in the passage? Let's go back to look for this one.**" (K3, T1P1Q13)
- [*Reads question and introductory sentence.*] "OK, summary. [*Reads all options and selects three.*] So, I already got three. Let me go back and check in the text." (C6, T2P1Q13)
- [*Reads and rereads question and introductory sentence.*] "**Here it asks what the passage mainly talks about.** [*Reads first three summary sentence options.*] **Obviously, this has been mentioned in the text, but S2 must be wrong, so I'll eliminate it. [***Continues to read and selects three options.*] **The whole passage mainly talks about factors and results of desertification. These points** [that the examinee chose] **are all main ideas, while the rest are either not mentioned or minor ideas. I'll go back** [to the passage] **and check my answers.**" (C2, T1P2Q13)
- [*Reads and rereads question.*] "So, this is a brief summary of the whole passage. I have to leave three of the options and pick three of them." (O3, T2P1Q13)
- "So, I'm reading the directions again and try to see what I'm not supposed to do. [*Rereads question.*] That means maybe some sentences are right, but that they are kind of minor ideas but not summary or introductory sentence." (C1, T1P2Q13)

2. Only one **reading** strategy emerged as being frequent enough to warrant reporting on it: *repeating/paraphrasing/translating words, phrases, or sentences immediately after reading them to improve understanding* (R9), which occurred at a moderate rate (.33). This strategy likely occurred at a lower rate in this item type than in the BC and I item

types because examinees were already fairly familiar with the passage by the time they reached this item, which was always last in the question set. Examples of strategy R9:

- [*Rereads second paragraph.*] "So, it's not kind of private form, but…just people have to, just one customer was allowed to view it at a time." (C1, T1P2Q13)
- [*Rereads last paragraph.*] "The legs were strong… The legs were certainly functional both on land and at sea." (C4, T2P3Q13)
- "I want to see the text. [*Returns to passage.*] Semiarid. …semiarid lands….Oh, it talks about severe dryness so the crop failures are common." (K2, T2P1Q13)

3. The remaining six test-taking strategies that occurred at a moderate rate or greater all related to how the respondents dealt with the options. These can be divided into two groups of strategies. The first set reflected how examinees went about deciding which options to select or discard. As with the Basic Comprehension and Inferencing item types, two strategies that occurred frequently were *using understanding of the paragraph/passage meaning to select* (T22) or *to discard* (T28) *specific options*, both occurring at very high rates (3.30 and 2.67, respectively) because examinees had to select and discard multiple options (sometimes changing their minds). The test designers had intended respondents to read in a more extensive fashion and to take into account the entire text in selecting possible options. When respondents were considering the overall passage meaning, they were being consistent with this testing aim. Examples of strategies T2 and T28:

    - [*Reads Option 1.*] "**This sentence is not a summary sentence.** [*Reads Options 2-6.*] **This sentence is obviously the main idea of the second and third paragraphs. Facial expressions are the external expressions on an individual's emotions, which may in turn influence the individuals expressions.**" (C8, T1P1Q13)
    - [*Reads Option 1.*] "This is true because the passage talked about increasing human population and how human population wants more food, so they'll be using more land for agricultural growth." (K2, T2P1Q13)

- [*Reads Option 4.*] "No, it doesn't talk about irrigation. It talks about excessive water, so this is not it." (K2, T2P1Q13)
- [*Reads Option 2.*] "I think that's one of the main ideas because in the passage the author explains that. [*Reads Option 3.*] No, there is nothing related to this in the passage, so it cannot be one of the main ideas." (O1, T1P1Q13)
- [*Reads Option 3.*] "No, this sentence seems incorrect; even the meaning is incorrect. [*Reads Option 5.*] This is true, and this is also a summary." (C3, T1P2Q13)

Because it generally co-occurred with strategies T22 and T28, the strategy *reconsidering or double-checking the response* (T19), which occurred at a high rate (.52), can be naturally grouped with them. Examples of strategies T19:

- [*Selects three options.*] "Let me read all the sentences in the box. [*Rereads Options 2, 3, and 4 that were dragged into box.*] OK, let me read the first option again. [*Rereads Option 1.*] This sounds correct, too, but I think it's more limited. I'll go with my initial picks." (O3, T2P1Q13)
- [*Reads Option 6.*] "**The passage did mention this. So that means my** [other] **selections were not correct.** [*Reviews three options already selected.*] **I'll drag out sentence 2 and drag in sentence 6.**" (J3, T2P3Q13)
- [*Rereads Option 2, which was already selected.*] "This is not right. [*Removes Option 2.*] So, I think this one should be right." [*Rereads and inserts Option 3.*] (C1, T1P2Q13)

4. The second set of strategies focusing on how subjects dealt with options all relate to how the examinees tried to make sense of the options: *considering and then postponing consideration of option(s)* (T16) occurred at a very high rate (2.93), *paraphrasing the meaning of the option* (T14) occurred at a high rate (.57), and *wrestling with option(s) meaning* (T17) occurred at a moderate rate (.37). Examples of strategies T14, T16, and T17:

    - [*Reads Option 1.*] "**This first sentence means that the two parties developed in the process of economic and political competition. This sentence seems to**

be the main idea of the whole passage, but I'm not quite sure so I'll put it aside for awhile." (C7, T1P2Q13)

- [*Reads Option 5.*] "Put it aside for awhile. [*Reads Option 6.*] **Let me have a look at the options again. Sentence 2 and 3 have already been eliminated. I need to go back to the text.**" (C7, T1P2Q13)
- [*Reads Option 1.*] "**I don't know about this. Move on to the next one.** [*Reads Options 2-4.*] **Not sure about this, but I think it was related to the northern part. Skip it.**" (K3, T1P2Q13)
- [*Reads Option 3.*] "I need to check if this one is correct or not. [*Continues with Options 4-6. Rereads Options 1-3.*] I'm still not sure about that. I'll come back for it. I'll go back to the passage to check." (O2, T1P2Q13)
- [*Reads Option 2.*] "Severe dryness becomes common. **Severe dryness is common**? As periods of severe dryness . . . **Severe dryness became common? I don't get this**." (J1, T2P1Q13)
- [*Reads Option 2.*] "**Thomas Edison developed… I don't understand this.**" (J3, T2P2Q13)
- [*Reads Option 2.*] ". . . **whale that lived both on the land and in the sea.**" (J3, T2P3Q13)
- [*Reads Option 1.*] "**Is this a summary? I'm not sure.** [*Reads Option 2.*] Facial expression **and** emotional states **interact with each other through a variety of** feedback**. I have to think more carefully about** facial emotion**, so I'll skip this.**" (J8, T1P1Q13)
- [*Reads Option 1.*] "…constituencies . . . What is the meaning of the last word? [*Rereads Option 1.*] OK, I think this sentence is hard for me. The meaning of this sentence is that these parties are developing in response to the needs, yes, the meaning of this sentence is that the two parties were developed as a response or like a reflection or feedback or influenced by the economic and political situation or development." (C3, T1P2Q13)

Reading to Learn-Schematic Table (R2L-st)

The Reading to Learn-schematic table (R2L-st) item type 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). Like the R2L-ps items, the R2l-st items were meant to call on respondents to *read through the entire text* in order to complete the table. For this R2L item format, respondents are to select five out of seven noun phrases and to drag them into a schematic table according to some organizing principle. The only example appearing in the *LanguEdge courseware* calls for identifying three constructive and two destructive forces with respect to soil erosion. The accuracy rate for the 10 attempts to respond to the 1 item of this type (five correct choices possible for the item) was 46/50 = 92% (J = 93%, C = 100%, K = 80%, O = 100%). Table 20 outlines the most frequently used reading and test-taking strategies for this item type.

Table 20

Common Strategies for Item Type R2L-st

Strategy code	Strategy description	Frequency rate
T22	Selects option(s) through vocabulary, sentence, paragraph, or passage overall meaning (depending on item type).	2.80
T16	Considers the option(s) and postpones consideration of the option(s).	1.90
T2	Goes back to the question for clarification: Paraphrases (or confirms) the question or task.	1.30
R9	Repeats, paraphrases, or translates words, phrases, or sentences—or summarizes paragraphs/passage—to aid or improve understanding.	1.20
R6	Reads a portion of the passage carefully.	1.00
T28	Discards option(s) based on vocabulary, sentence, paragraph, or passage overall meaning as well as discourse structure (depending on item type).	1.00

(Table continues)

Table 20 (continued)

Strategy code	Strategy description	Frequency rate
T4	Reads the question and considers the option(s) before going back to the passage/portion.	.90
R7	Reads a portion of the passage rapidly looking for specific information.	.70
T1	Goes back to the question for clarification: Rereads the question.	.50
T17	Considers the option(s) and wrestles with option(s) meaning.	.50
T19	Reconsiders or double-checks the response.	.40

Note. Strategies are ranked and grouped according to three levels of frequency rate(number of occurrences / number of items of that type): very high (≥ 1.0), high (≥ .50), and moderate (≥ .30).

In reviewing the reading and test-taking strategies that occurred at very high (≥ 1.0), high (≥ .50), and moderate (≥ .30) rates for this item type, as given in Table 20, the following notable strategy trends emerged:

1. As with the R2L-prose summary items, the strategy of *reading the option(s) before going back to the* passage (T4) occurred at a high rate (.90). Again, this strategy was particularly common because examinees apparently felt they had a good handle on the main ideas in the passage from working through the previous items on the test. As noted with the R2L-ps items, this runs counter to how examinees approached the BC and I item-types, for which the more common strategy was to return to the passage first before considering the options. Two other strategies cluster with this one in that they all focus on how subjects dealt with the question itself: *paraphrasing (or confirming) the question or task* (T2), occurring at a high rate (1.30), and *rereading the question for clarification* (T1), which occurred at a moderate rate (.50). Examples of strategies T4, T1, and T2:

    - [*Reads question.*] "OK. Destructive and constructive? Interesting. I'm going to read the options first." (O7, T1P3Q12)
    - [*Reads question.*] "**What's the meaning of** destructive **and** constructive process?" [*Rereads question.*] (J2, T1P3Q12)

- [*Reads question.*] "OK, I need to find the constructive and destructive processes of erosion. Let me read the options first. [*Reads the options.*] OK, I'll look at the passage and find these constructive and destructive processes." (O2, T1P3Q12)
- "**Constructive means making something. Destructive means** . . ." (K1, T1P3Q12)
- "**Seems we need to choose specific examples.** [*Reads Option 1.*] **This one should be one of them, which is based on the theory of drifting lands hitting each other.**" [*Continues reading following options.*] (C7, T1P3Q12)
- "**This asks you about three factors which lead to the formation of mountains. This second category asks about destructive forces.** [*Reads options*.] **Yes, the weather should be one of them. Right, wind-driven sand should be another one.**" (C5, T1P3Q12)

2. Three reading strategies used by examinees while working through this item type emerged frequently enough to warrant reporting on them. *Repeating/paraphrasing/translating words, phrases, or sentences immediately after reading them to improve understanding* (R9) and *reading a portion of the passage carefully* (R6) occurred at very high rates (1.20 and 1.00, respectively), and *(re)reading a portion of the passage rapidly looking for specific information* (R7) occurred at a high rate (.50). This task does appear to have activated the use of reading strategies in order to perform the required organization of phrases. Examples of strategies R9, R6, and R7:

    - [*Reads Paragraph 3.*] "**Earthquakes and volcano activity are** uplift. **They are** constructive." (J2, T1P3Q12)
    - [*Rereads first three paragraphs.*] "In this paragraph it says the formation of mountains began with the phase of crashing into each other as a constructive process. [*Continues reading passage.*] Here it says earthquakes also may result in the foundation of mountains. But I'm not sure if it's constructive or destructive., so I'll continue to read. [*Continues reading passage.*] Right here it talks about destructive forces." (O2, T1P3Q12)
    - [*Rereads passage.*] "**This paragraph talks about** mountains. **And this paragraph talks about** volcanoes. [*Rereads next paragraph.*] **The story of**

mountain **is** earthquake. Earthquake **forms…What? Let's go back.**" (J8, T1P3Q12)

- "**Well, I'm not supposed to insert all choices in the box. Let's go back to the passage**." [*Skims the entire passage*.] (J5, T1P3Q12)
- [*Reads first paragraph*.] "Constructive process **is**…uplift."(J2, T1P3Q12)

3. Again, in the same vein at the R2L-prose summary items, the remaining six test-taking strategies that occurred at a moderate rate or greater for the R2L-st items all dealt with how examinees dealt with the options. These can be divided into two groups of strategies. The first set reflects how examinees went about deciding which options to select or discard. As with the Basic Comprehension and Inferencing item-types, *using understanding of the paragraph/passage meaning to select* (T22) or *to discard* (T28) *specific options* were two strategies that occurred frequently, both occurring at very high rates (2.80 and 1.00, respectively) because examinees had to select and discard multiple options (sometimes changing their minds). The test designers had intended respondents to read in a more extensive fashion and to take into account the entire text in selecting possible options. So, just as in the case of the R2L-prose summary item type, respondents **did** appear to be reading in a more extensive fashion and to be taking into account the entire text in selecting possible options. When respondents were considering the overall passage meaning, they were being consistent with this testing aim. Examples of strategies T22 and T28:

    - [*Reads Option 3.*] "**The passage said sand prevents rock erosion. So this is not right . . .** [*Reads Option 7.*] **Yes, the passage said acid rain causes** erosion, **so** weather process **is destructive** . . . . **The passage did not mention Sentence 2 and Sentence 4, so they are not right.**" (J2, T1P3Q12)
    - "Based on the paragraph, earthquakes are also constructive. Again, based on the paragraph, the option 'wind-driven sand' is also destructive." (O2, T1P3Q12)
    - "**The grass roots, on the contrary, can result in prevention of erosion. In other words, it shouldn't be among the destructive forces**." (C7, T1P3Q12)
    - "The paragraph says the formation of the mountains began with the crashing into each other as a constructive process, so the first choice 'collision of Earth's crustal plates' is a constructive process." (O2, T1P3Q12)

- [*Reads Option 5.*] "**This occurred but it is not** constructive." (T1P3Q12)

Because it generally co-occurred with strategies T22 and T28, the strategy *reconsidering or double-checking the response* (T19), which occurred at a moderate rate (.40), can be grouped together with the other two. Examples of strategies T19:

- [*Rereads option choices.*] "**Oh, sentence three goes here**." [*Moves it from constructive to destructive side of table.*] (J8, T1P3Q12)
- "**Well, something is wrong here. I see, sentence 1 is** constructive." [*Moves sentence from destructive to constructive side of table*.] (J5, T1P3Q12)
- [*Skims passage after making initial selection of choices*.] "**Now I understand the definitions** [of constructive and destructive]**. I should drag out all of my answers from the boxes** [and start again.]" (J5, T1P3Q12)
- "I believe the option 'volcanic activity' is correct because they are related to mountains, but let me go back to that portion of the passage to make sure." (O2, T1P3Q12)

The second set of strategies focusing on how subjects dealt with options can be grouped by how examinees worked to make sense of the options: *postponing consideration of option(s)* (T16) occurred at a very high rate (1.90) and *wrestling with option(s) meaning* (T17) occurred at a high rate (.50). Examples of strategies T16 and T17:

- [*Reads Option 1.*] "**What is the meaning of** collision?" (J2, T1P3Q12)
- [*Reads all the options.*] "**Well, I'm not sure. I'm going to back to the passage.**" (K4, T1P3Q12)
- "**Grass roots should be helpful for soil rather than destructive; I don't understand this.**" (C5, T1P3Q12)
- [*Reads Option 4.*] "**Put it aside.**" (C7, T1P3Q12)
- [*Reads Option 6.*] "Volcanic activity. [*Leaves it and goes to last option.*] **This is here** [constructive]."
- [*Reads Option 2.*] "**Let's go back to the passage…**separation of the continent. [*Skims passage.*] **I see** Option 7 weather process **here.**" [*Selects and moves on the Option 6.*]. (J5, T1P3Q12)

Comparison of the Inferencing and Reading to Learn Items With the More Traditional Basic Comprehension Items

This section will respond to the second research question dealing with whether the Inferencing and the Reading to Learn items require and evaluate different and supposedly more difficult approaches to academic reading than the Basic Comprehension items. We found this **not** to be the case. In fact, it could be argued that all of the item types assessed academic reading skills in similar ways. In other words, we did not find through our close analysis of strategy use that the two new types of items prompted noticeable use of those text processing strategies associated with having to cope with longer academic texts—such as more elaborate strategies for retaining ideas in working memory, refined strategies for perceiving the gist, more finely-tuned strategies for identifying logical connectors and determining how sections of passages interrelate in an effort to establish passage coherence, or enhanced strategies for management of details. First, we will look at the level of focus, second at the strategies deployed, and third at the difficulty of the item types.

The Level of Focus Within the Text

Where differences in strategy use arose across item types, it was often a matter of where respondents needed to look for the information in order to answer the given item type. In most cases they were **directed** to look for information at the sentence, paragraph, or passage level, but in some cases, they went to the passage level by default due to a lack of direction about where to locate the information. The Reading to Learn items were, by design, intended to measure examinee's overall understanding of the ideas and organization of the complete test passage, and this was generally found to be the case. With regard to the other item formats, the level of focus was found to vary, and consequently, how examinees approached these items varied even within the broader categories of the Basic Comprehension and Inferencing item types.

In effort to describe robustly the actual focus of each of the item types, let us start with a statistical tally of which item types had a sentence-, paragraph-, or passage-level of focus:

- BC-v: 100% (20/20) were focused on the sentence level, but some of them also gave clues to the reader in sentences that preceded or followed the sentence the targeted word was in.
- BC-f: 77% (17/22) were focused on the paragraph level and 23% (5/22) started with "according to the passage…" or the like.

- BC-n/e: 67% (2/3) were focused on the paragraph level and 33% (1/3) required looking over the whole passage to respond.
- BC-ss: 100% (6/6) were at the sentence level, although the context of surrounding sentences would help.
- BC-pr: 100% (4/4) were at the paragraph level in that readers had to look for the pronoun referent in a sentence other than the one the pronoun was in.
- I: 67% (4/6) were at the paragraph level and 33% (2/4) were at the passage level.
- I-it: 100% (6/6) were at the paragraph level, but several had possible slots for inserting the test sentence at the start of a paragraph, which would require readers to have an understanding of the ideas in the previous paragraph.
- I-rp: 100% (3/3) were at the paragraph level.
- R2L-ps: 100% (5/5) were at the passage level
- R2L-st: 100% (1/1) was at the passage level

Now let us take a more in-depth look at the focus of a few of the items. For example, while two of the three Basic Comprehension–not/except item types focused on the paragraph level (with the stem following this format: "According to paragraph 6 . . . "), one of the three BC-n/e items required examinees to consider the whole passage. Again, this proved to be a much more difficult task. The stem for this item read: "All of the following are mentioned in the passage . . ." (T2P1Q9). Similarly, most of the Basic Comprehension-factual information items directed respondents to specific paragraphs, but some did not, requiring respondents to draw on their previous understanding of the text structure to locate the relevant information, or in the case of some, to skim or scan from the start of the passage to find the relevant information. Specifically, while 78% (18/23) of the BC-f item types focused on the paragraph level, (with the stem generally following this format: "According to paragraph 4,…" or otherwise pointing to a specific spot in the passage), 22% (5/23) of the items required examinees to consider/search through the whole passage. The stem for these questions generally started: "According to the passage, . . . ," or otherwise did not point to a specific paragraph. This difference was found to make these items more challenging.

With regard to the Basic Inference (I) item type, while four of the six items focused on the paragraph-level, (with the stem generally following this format: "Which of the following can be inferred from paragraph 1 . . ."), the other two items had a stem that did not point readers to a specific paragraph, requiring readers to work with the whole passage to determine the answer.

The Reading and Test-Taking Strategies Deployed

While there were strategies used just for certain item types and not for others, there were two reading strategies and six test-management strategies that tended to be used in responding to the full range of item types (see Table 7).

The reading strategies were: *reading a portion of the passage carefully* (except for R2L-ps) and *repeating, paraphrasing, or translating words, phrases, or sentences—or summarizing paragraphs/passage—to aid or improve understanding*. Test-management strategies that were often used across items include:

- 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* (except for BC-v and BC-pr)
- T5—*reads the question and then reads the passage/portion to look for clues to the answer either before or while considering options* (except for R2L-ps and R2L-st)
- T16—*considers the options and postpones consideration of the option* (except for I-it)
- T22—*selects options through vocabulary, sentence, paragraph, or passage overall meaning*
- T28—*discards options based on vocabulary, sentence, paragraph, or passage overall meaning as well as discourse structure*.

All of these strategies reflect the fact that respondents were in actuality engaged with the reading test tasks in the manner desired by the test designers. The consistent and frequent use of the above strategies shows that respondents were actively working to understand the text, to understand the expectations of the questions, to understand the meaning and implications of the different options in light of the text, and to select and discard options based on what they understood about the text.

What set the Reading to Learn item types off from the other item types was that the respondents reported focusing more on passage-level understanding than with the other two sets of item types. For the Inferencing item types the focus was more on the paragraph level, as was the case with the Basic Comprehension item types, with the exception of BC-v—where respondents were focused more on the word/sentence level. The problem with making generalizations is that they do not necessarily hold. For example, there were several cases of BC items calling for passage-level processing, such as one BC-n/e item where the focus was found to be more on the passage level.

The Difficulty Level of the Items as Determined by Success Rate

Before reporting on item difficulty, a few caveats are in order. This study was intended largely as a qualitative exploration of strategy use among a limited number of respondents rather than with a representative TOEFL population, and not all the LanguEdge items for each type were included. In addition, the innovative nature of the R2L items also meant that calculation of results on them was more complex than with the other items, since partial credit items have thresholds for credit that are not comparable to single selection multiple-choice items.[11] In addition, the actual item difficulty would be a function of both the respondents' language proficiency (and in this case, it was high), as well as by sometimes subtle variations in how the actual item or task is configured (i.e., the wording of the question, features in the passage, and so forth). In essence, our study results need to be interpreted in their rightful context as representing results with an advanced group of respondents, dealing with a subset of LanguEdge reading items that were deemed by ETS to be representative of both the content and difficulty level of what can be expected to appear on the new TOEFL. Thus, having these caveats in mind, our results would simply constitute food for thought in that our close-order analysis yielded rich descriptive data of just what made certain items easier or more difficult for these respondents and why. Usually, analysis of difficulty is performed strictly through number crunching, and reasons for difficulty are simply intuited by the researchers.

In our small-scale study, the R2L item types were not necessarily found to be more difficult for respondents even though they did require a more passage-level processing. Many respondents struggled more with the BC-sentence simplification and BC-factual information types, for example, than with the R2L-prose summary items. Similarly, the Inferencing item types (like I-insert text) were not necessarily as challenging as some of the BC item types either. In fact, the

two R2L item types and the I-insert text item type actually had the **highest** percentages of correct responses (91-92%). The Basic Inference item type was by far the most difficult (56% right), but three BC item types (BC-v, BC-ss, and BC-f) were also somewhat difficult (with an 81% success level). It turned out that on the average, the three Inferencing item types proved most challenging (77% success rate, although this was due mostly to the extreme difficulty of the Basic Inference item type), the five BC items the next most challenging (83%), and the R2L items the easiest (90%), if we take the average percent correct within each broad type. Although not so difficult comparatively, the R2L items did take longer, in general, for subjects to complete because they focused on passage-level understanding and required more steps in the response process.

The Difficulty Level as Suggested by Verbal Report Data

The report by respondents of using strategies T3 (*wrestling with question intent*) and T17 (*wrestling with option meaning*) were considered indicators of whether they found particular item types (and specific questions) to be more challenging. The use of strategy T3 was only reported at a measurable frequency for one item type, namely, the Basic Inference (I) item type, while strategy T17 was at a measurable frequency for not only an Inferencing item type (I-rp), but also for a Basic Comprehension item type (BC-ss), as well as for both of the R2L item types. Similar to T17, T19 (*reconsidering or double-checking the response*) was found in measurable reported use for two Basic Comprehension item types (BC-ss and BC-n/e) and for the Reading to Learn-prose summary format. Other indicators that an item was challenging could be T12 (*selecting preliminary options*), which was measurable as a strategy in four of the five BC item types and for two of the three Inferencing formats and T16 (*postponing consideration of an option*), which was at a measurable level for all item types except for the insert text type of Inferencing items (I-it).

It could be argued that academic reading skills are being brought to bear on some of the questions and options themselves, as examinees work to integrate what they know from the text with what they are expected to do with the test item. This may not be inappropriate, but test designers should be careful to insure that test items are written in such a way that they are not a major source of confusion. For example, Question 9 of Test 1, Passage 2, seemed to be particularly difficult just for the examinees to understand—no doubt partly because it contains seven prepositional phrases ("Which of the following can be inferred from paragraph 5 about differences in the strength of political beliefs of groups within the Whig Party?").

Discussion

The underlying goal of this study was to gain a better understanding of how reading and test-taking strategies are used on tests as part of the process of construct validation (Anderson et al., 1991). The focus was on the strategies that the respondents used in producing answers to the five Basic Comprehension item types, the three Inferencing item types, and the two Reading to Learn item types on the reading section of the LanguEdge tests, designed to familiarize respondents with the new TOEFL.[12] The basic assumption being made in this study is that the way respondents deal with testing tasks on the LanguEdge tests is similar to the way that they would react to reading tasks on the new TOEFL itself.

Of particular interest to us was whether the Reading to Learn and the Inferencing items required and assessed different approaches to academic reading from those elicited by the Basic Comprehension questions. This appeared not to be the case on the basis of our study of strategy use. In addition, we were interested in whether these innovatively different tasks would also be more difficult. In their statement of the design principles for the new TOEFL, Enright et al. (2000, p. 6) stated that while "easy tasks could be designed for reading to learn . . . and difficult tasks asking examinees to find discrete information or read for general comprehension could be designed by manipulating task and linguistic/syntactic variables . . . [s]till, we are more likely to find reading to learn . . . associated with more challenging academic tasks and to require more sophisticated processing abilities than reading to find discrete information or reading for general comprehension." The accuracy rate would suggest that whereas some of the Inferencing items were among the most difficult for the respondents, the Reading to Learn items were among the easiest. So our findings are not consistent with the expectations of the TOEFL committee based on the design principles document.

We used verbal report methods in this endeavor. By asking the test-takers to think-aloud as they worked through these various item types, it was possible to analyze the resulting protocol to identify the cognitive processes involved in carrying out the tasks. The intent was to gather concurrent verbal report (i.e., a description of what they were doing while they were doing it) in order to get an impression of how the readers processed the texts in an effort to answer the questions.

The first general finding was that subjects approached the new TOEFL reading section as a test-taking task that required that they perform reading tasks in order to complete them. In other

words, the primary goal of the subjects was to get the answers right, not to learn or gain anything from the texts read. The second finding was that the strategies deployed were generally consistent with TOEFL's claims that the successful completion of this test section requires academic reading-like skills. Overall, the findings of the study supports the statement made by ETS with regard to the new TOEFL that it requires examinees to have both a local and general understanding of the test passages. We found that the respondents in our study did in fact tend to draw on their understanding and interpretation of the passage to answer the questions, except when responding to certain item formats like Basic Comprehension-vocabulary, where many subjects answered from their own background knowledge if they were already familiar with the targeted word.

The Basic Comprehension Item Types

The Basic Comprehension items were seen to require academic reading skills as applied to a testing task as follows:

1. They minimally required understanding at the sentence level, but in most cases the paragraph level, and even in a couple of cases at the passage level.
2. They challenged the respondents to use powers of inference when they did not recognize vocabulary.
3. They called for being mindful of cohesion and coherence issues when attempting to fit the portion of text into the larger discourse.
4. Because the texts were so long, the respondents had to draw more on their memory of what they had read, as well as on what they had learned about the passage from responding to previous questions.
5. Consistent with the previous point, respondents were "learning" as they were doing the test; they gained information from one set of questions, which they then were able to apply to the next set of questions.
6. While ostensibly a lot of the items were very local, the fact that the text was large meant that respondents needed more than a local reading.
7. The respondents were clearly engaged in problem-solving activities.

8. The length of the texts and the nature of the items seemed to preclude the use of test-wiseness strategies for the most part, thus making the tasks more than a matter of completing testing tasks by circumventing the need to exercise their actual language knowledge or lack of it.

The Basic Comprehension-vocabulary, the Basic Comprehension-factual information and the Basic Comprehension-sentence simplification items proved to be among the most challenging item types on the test. It is difficult to pinpoint precise reasons for this. For the BC-vocabulary items, it may well be due to the fact that many examinees relied heavily on their background knowledge—which may not have been an asset—when answering these rather than making a focused effort to be sure their understanding of the word made sense in the context of the sentence. For the BC-factual information items, examinees had to work with the text on at least a paragraph level and in several cases at a passage level in order to answer the question. Consequently, there was often much more text to work with than has traditionally been the case with such items on the TOEFL, which as noted previously is one means for increasing the challenge of test items. The BC-sentence simplification items no doubt proved difficult because they required examinees to "transform" their understanding of a sentence into a synonymous form, a task that is made more challenging as it generally required that they understand the meaning of the original sentence within the larger context of the paragraph. This is a fairly difficult task.

With regard to the Basic Comprehension-vocabulary items, we would offer an explanation for why certain common strategies for dealing with such items were underutilized. As reported above, we found only a low rate of occurrence of the two test-management strategies *picking an option mainly because the other options were unreasonable based on background knowledge* (T23) or *based on paragraph/overall passage meaning* (T24) and of the test-wiseness strategy *using 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* (TW1). Likely explanations for this finding would be the high proficiency level of the respondents, as well as being an artifact of the research situation where these paid subjects felt obliged to do more than simply use the process of elimination for selecting an option. Perhaps in a real testing situation, where they were not being observed, they would use these strategies more.

The Inferencing Item Types

The Inferencing items were seen to require academic reading skills as applied to a testing task as follows:

1. They required understanding of the text at the paragraph level, and in many cases at the passage level. The data revealed significant efforts on the part of respondents to gain this understanding while working with these item types, including rereading, reading purposefully and carefully, and paraphrasing/translating to facilitate comprehension.
2. They indeed challenged the respondents to use powers of inference to recognize (a) an argument or idea that was strongly implied but not explicitly mentioned, (b) the rhetorical purpose, as well as (c) lexical, grammatical and logical links in order to determine where to best insert a new sentence. Related to the first point above, respondents clearly drew on their understanding of overall paragraph and passage meaning and structure to consider and evaluate, and then select or discard options in order to complete the Inferencing item tasks.
3. Particularly for the I-it, but also to some extent the Basic Inference (I) item types, the Inferencing items required attention to markers of cohesion, a sense of coherence and textual clues in order to correctly respond to the items.
4. The length of the texts required respondents to draw more on their memory of what they had read, as well as on what they had learned about the passage from responding to previous questions.
5. Consistent with the previous point and as observed with the BC items, respondents were learning as they were doing the test; they gained information from one set of questions, which they then were able to apply to the next set of questions.
6. The respondents were clearly engaged in problem-solving activities. Not only did they make efforts to gain a clear understanding of the text, but they also worked to understand and evaluate the test question itself and the options that were given in light of the text.
7. As observed with the BC items, the length of the texts and the nature of the items seemed to preclude the use of test-wiseness strategies for the most part, thus making the

tasks more than a matter of completing testing tasks by circumventing the need to exercise their actual language knowledge or lack of it.

As noted previously, the Basic Inference (I) items proved to be the most difficult items overall with only a 77% overall success rate. In particular, the two Basic Inference items that required a passage-level orientation (because no specific paragraph was referenced in the item as was the case for the others) together had an even lower success rate, 65%. The verbal reports from these items indicated that not only did the task require a high level of understanding of the text, but that examinees also had a harder time understanding the meaning and intent of the question itself than with other questions. The Basic Inference (I) item type appears to be a strong discriminator of academic reading ability, but test designers do need to be cautious that the question itself is well written and not the source of confusion and error.

We originally had our doubts about whether the Inferencing items aimed at assessing the respondents' ability to identify the rhetorical purpose (I-rp) would truly elicit inferencing behavior. In fact, the I-rp items were found to engage the respondents in inferencing activities, consistent with the goals of the test constructors. Our analysis of the strategies used to find responses to these items demonstrated that the respondents were deriving their answers through the kind of reasoning procedures that are called upon in an academic context.

The Inferencing-insert text items that require examinees to insert a new sentence into the most appropriate of four places in a paragraph were among the easiest of the item types. Most notably, this item type drew specifically on examinees' understanding of the discourse structure, other grammatical structural features (use of pronouns for nouns, etc.), and lexical and logical links within a paragraph more than other item types, as might be expected.

The Reading to Learn Item Types

According to the test constructors for the new TOEFL, the truly innovative formats focused on the "reader purpose perspective" in the test design and are associated with the academic purpose of reading to learn when faced with comprehending longer reading passages. One of the requirements of being able to successfully read to learn is the ability to "integrate and connect the detailed information provided by the author into a coherent whole" (Enright, et al., 2000, p. 6). One of the Reading to Learn formats intends to measure the extent to which L2 readers can complete a

prose summary through questions that are referred to as multiple-selection multiple-choice responses. It entails the dragging and dropping of the best descriptive statements about a text.

A question we asked ourselves as researchers was, "To what extent does this kind of task simulate the processes that respondents go through when identifying (or stating to themselves) the points that could contribute to summarizing a text?"

First of all, it has become readily apparent that summarizing is not ultimately a strategy, nor are skimming and scanning (see, for example, Uruquhart & Weir, 1998, pp. 96-98). They are reading skills that may be operationalized by means of a series of strategies. To summarize successfully, respondents need both reading and writing skills. First, they must select and use effectively those reading strategies that are appropriate for summarizing the source text—that is, identifying topical information, distinguishing superordinate from subordinate material, and identifying redundant as well as trivial information. Then they must perform the appropriate writing tasks to produce a coherent text summary—that is, selecting topical information or generating it if none appears explicitly in the text, deleting trivial and redundant material, substituting superordinate terms for lists of terms or sequences of events, and, finally, restating the text so that it sounds coherent and polishing it so that it reads smoothly (Kintsch & van Dijk, 1978; Brown & Day, 1983; Chou Hare & Borchardt, 1984).

In principle, a summarization testing task is intended to reflect how respondents might perform on such a task in an authentic setting. Test summaries such as the type of abbreviated summary-like task included in the new TOEFL as Reading to Learn-constructed response and Reading to Learn-schematic table usually have been found to support comprehension and learning of textual material (Grabe, 2004). In our study, in order to perform these tasks, the respondents were found to need at least some understanding of the rhetorical pattern of the passage (e.g., perceive that the text classifies types of soil erosion), and a working knowledge of how to select the key points from a larger set. Still, the tasks do not approximate real-life summarization tasks for various reasons. For one thing, the current approach does not call for removing the text from the screen (and allows respondents to access it whenever they wish), the text is always available to help jog the reader's memory and to give clues. Contrast this with real-life tasks where often the source text is not available at the time the summary is being written.[13]

Since in the LanguEdge measure, no writing is called for in the Reading section, and even the set of possible main points is provided for the respondents so that they only need to select those

that they are to drag into a box (whether astutely or by guessing), they do not need to find main statements in the text nor generate them (e.g., by reconceptualizing lower-level ideas at a higher level of abstraction). Where strategies are called for on the Prose Summary item type is in distinguishing the superordinate statements from the subordinate, usually more detailed ones. But even this process of distinguishing superordinate from subordinate ideas Cohen (1994b) found to be a challenging task for Brazilians reading texts and preparing summaries in English as a foreign language, primarily because they had insufficient grasp of the vocabulary to determine what was central and what was more secondary in nature. The question is just how strategic the readers have to be to perceive which statements are more central and which more secondary. In the current research, the respondents were advanced enough in their English proficiency so that they handled the statements with ease (91-92% success rate).

In sum, both of the R2L item types clearly required a broad understanding of the major ideas in the passages and the relative importance of them, in keeping with the intended purpose of this item format, namely, to reflect the respondents' ability to read and process longer texts effectively. But we need to bear in mind that the completion of these items was greatly facilitated by the fact that they were always the last items in the test and so examinees had already read the passage—at least significant parts of it—several times in order to answer the preceding 11 to 12 items. Indeed, many began considering and selecting from the options even before returning to the text because of their familiarity with the text by that point in the test. It seems examinees found the R2L items relatively easy for two reasons: (1) they had become quite familiar with the passage and the key ideas because of their efforts to answer all the other items that always came before them; and (2) examinees did not have to generate their own summary statements of the key ideas, merely select those that had been prepared for them.

Hence, whereas the aim may have been to construct academically more demanding items (e.g., requiring strategies for retaining ideas in working memory, for identifying logical connectors and other markers of cohesion, for determining how sections of passages interrelate in an effort to establish passage coherence, for perceiving the gist, and for managing details), the reality was that the R2L items were less demanding than they probably would have been had they appeared as the sole items accompanying a lengthy text. Actually, this finding exposes a paradox about test construction. The TOEFL committee's rationale for placing the Reading to Learn items at the end of each item set was to allow examinees to prepare for this task by completing other items in an

item set. Thus, our finding that the participants were already quite familiar with the text by the time they reached these items confirmed this design rationale, but also meant that the range of what was being tested in these items was somewhat constricted due to repeated prior exposure to the text.

An example from one of the Japanese subjects illustrates the ease that some of our subjects had in responding to the R2L-prose summary item and the lack of need many felt to return to the text to look for clues about choices or to confirm their selections:

> [*Reads item directions.*] "**So I should choose three** summaries." [*Reads introductory sentence.*] [*Reads first answer choice.*] "**This seems right.**" [*Reads second answer choice.*] "**This seems wrong. I think the first one is correct so I should choose it.**" [*Drags first answer choice to box.*] [*Reads third answer choice.*] "**This is correct, too.**" [*Drags third answer choice to box.*] [*Reads fourth answer choice.*] "**This is also mentioned.**" [*Reads fifth answer choice.*] "**This is wrong. The fourth one is correct.**" [*Drags fourth answer choice to box.*] [*Reads sixth answer choice.*] "**I am fine with the three that I have selected. The others are mentioned, but they are not important.**" [*Note: All choices were correct.*] (J7, T2P1Q13)

While many subjects worked a little harder than this Japanese subject to respond to the R2L items, it should be reiterated that in fact very little "whole passage" processing was occurring while subjects were working through these items. Subjects were not looking at the reading fresh, summarizing in their heads what the key ideas were and the text organization, and then moving to the test item to find the answer that matched the understanding they had in their heads. If one looks at the strategies that occurred at the highest rates for the R2L items, they almost all deal with examining the options one by one and selecting or discarding them. As we noted, subjects are approaching this as a testing task, not a reading task, which explains why the strategies we are seeing are overwhelmingly test-taking strategies, and ones that focus on specific options.

Limitations

Since this was essentially a qualitative analysis, with some effort made to quantify the verbal report data for the purpose of comparing strategies for frequency of use, we must be wary of putting too much faith in the strategy frequency counts. At best, the data indicate trends in strategy use as reported by the respondents or clearly observable to the RAs who were coding the

strategies. Despite our best efforts to ensure that the four RAs coded the verbal report, as well as their observed and preverbalized behavior, in a consistent fashion, it is still possible that there were inconsistencies in the interpretation of the strategy categories.

It should also be noted that there undoubtedly were other strategies in use that were **not** described in the verbal reports. In addition there could have been strategies appearing in the verbal report that were actually used more or less than indicated. Furthermore, the fact that subjects were verbalizing as they worked through the items probably had some influence on how they went about completing the task. It is impossible to eliminate the reactive effects of verbal report on task performance.

As suggested above, it is also possible that respondents were making an effort to respond to each item more conscientiously than they would under normal circumstances. For one thing, the completion of the test was not timed, as it is under normal circumstances. For another thing, this was a low-stakes task since their score was of no consequence to the respondents. So in that sense, the conditions were different from those in place when respondents actually take the test. Furthermore, these were relatively high-proficiency students, and in addition they were either from East Asia (from China, Japan, and Korea) or from Turkey, the Middle East, or from other parts of Asia. Consequently, we need to be careful about generalizing these findings to students at other ESL proficiency levels and from other parts of the world.

Furthermore, a distinction was not made between strategies used for test items that were answered correctly as opposed to those answered incorrectly. A closer look at this variable might provide us with an even clearer picture regarding the test takers' effectiveness of their strategy use.

Nevertheless, we feel that the data are clearly reflective of the kinds of strategic moves that respondents do make in an effort to answer the different types of reading items on the new TOEFL. Consequently, we see the data as helpful in evaluating the item types to see what examinees, in general, are doing to answer the questions.

Further Analyses of the Data From This Study

There are still analyses that can be performed on these data, which do not respond to the research questions we posed, but which would nonetheless be of interest. For example, in the current report, we did not make any effort to distinguish respondents' strategy choices or success on the test by language group, partly because we were reluctant to make generalizations based on a maximum of 8 respondents in a group. Furthermore, the main differences seemed to be primarily

due to proficiency level, not to language group. Still, it would be possible to revisit the data with an eye to determining if any language- and culture-specific trends do emerge for the Chinese, Japanese, Korean, and Turkish students.

Another analysis that will be conducted on these data is intended to explore the value of a sociocultural perspective to explain how L2 readers/test-takers use their language resources to help them pursue intrapersonal, problem-solving tasks. While it is generally the case that L2 readers actively engage their expertise in their L1 to mediate comprehension, the question arises as to whether we can we make generalizations about when and why L2 learners draw on their L1 when engaged in L2 problem-solving activities. Since respondents were given an option to respond in the L1 or L2 during their verbal report in this study, it is possible to return to these data to see if there is any identifiable pattern to the use of one language or another in this process. While it is reasonable to assume that the choice of giving verbal report in L1 may be related to L2 language proficiency, we would want to see if the use of the L1 or L2 in verbal reporting is also associated with the use of a given reading or test-taking strategy.

Conclusion

This study set out to determine whether the new TOEFL is actually measuring what it purports to measure, as revealed through verbal reports. A test claiming to evaluate **academic** reading ability would be expected to include tasks calling for test takers to actually *use* academic reading skills in responding to items, rather than being able to rely on test-wiseness tricks. It was our finding that as a whole the Reading section of the new TOEFL does, in fact, require examinees to use academic reading skills to gain both a local and general understanding of the test passages—at least for respondents whose language proficiency is sufficiently advanced so that they not only take the test successfully, but can also tell us how they do it as well.

Nevertheless, it was also clear that subjects approached the new TOEFL reading section as a *test-taking task* that required that they perform reading tasks in order to complete it. In other words, the primary goal of the subjects was to get the answers right, not to necessarily learn, use or gain anything from the texts read. Thus, for these respondents, working their way through the Reading sections of the LanguEdge test did not *truly* constitute an academic reading task, but rather a test-taking task with academic-like aspects to it. While the respondents were found to use an array of test-taking strategies, they were primarily test-management strategies, and not test-

wiseness strategies. Also, they were perhaps reluctant to use test-wiseness strategies because they knew we were observing their behavior closely.

The second issue explored in this study was whether the Reading to Learn and the Inferencing items required and evaluated different academic reading skills from those elicited by the Basic Comprehension items. Overall, this was found not to be the case. Based on the findings from this primarily qualitative study, we would contend that *all* of the item types on the LanguEdge prototypical tests, including the Basic Comprehension items, in fact assess similar components of academic reading, as well as test-taking ability. The modifications to the Basic Comprehension item types, by placing them into a larger context that requires examinees to consider words and sentences in the context of larger chunks of text and even whole passages, have made them reflect academic-like tasks that elicit strategies comparable to those required of the Inferencing and Reading to Learn tasks. While the tasks and expectations for the three broad item types—Basic Comprehension, Inferencing, and Reading to Learn—are clearly different, they all tend to draw on the same sorts of strategies from respondents. So, in conclusion, the new TOEFL is assessing the ability of examinees to use a fairly consistent set of basic academic reading and test-taking skills to accomplish a variety of academic-like reading and test-taking tasks.

References

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.

Anderson, N. J. (1991). Individual differences in strategy use in second language reading and testing. *Modern Language Journal*, *75*(4), 460-472.

Anderson, N. J., 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.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, England: Oxford University Press.

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

Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. London: Longman.

Bernhardt, E. (1991). *Reading development in a second language: Theoretical research and classroom perspectives*. Norwood, NJ: Ablex.

Birch, B. M. (2002). *English L2 reading: Getting to the bottom*. Mahwah, NJ: Lawrence Erlbaum.

Block, E. (1986). The comprehension strategies of second language readers. *TESOL Quarterly*, *20*(3), 463-494.

Brown, A. (1993). The role of test taker feedback in the test development process: Test takers' reactions to a tape-mediated test of proficiency in spoken Japanese. *Language Testing, 10*(3), 277-303.

Brown, A. L., & Day, J. D. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior, 22*(1), 1-14.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1-47.

Carrell, P. L., & Grabe, W. (2002). Reading. In N. Schmitt (Ed.), *An introduction to applied linguistics* (pp. 233-250). London: Arnold.

Cavalcanti, M. C. (1987). Investigating FL reading performance through pause protocols. In C. Faerch & G. Kasper (Eds.), *Introspection in second language research* (pp. 230-250). Clevedon, England: Multilingual Matters.

Chou Hare, V., & Borchardt, K. M. (1984). Direct instruction of summarization skills. *Reading Research Quarterly, 20*(1), 62-78.

Cohen, A. (1984). On taking language tests: What the students report. *Language Testing, 1*(1), 70-81.

Cohen, A. D. (1986). Mentalistic measures in reading strategy research: Some recent findings. *The ESP Journal*, *5*(2), 131-145.

Cohen, A. D. (1987). Using verbal reports in research on language learning. In C. Faerch & G. Kasper (Eds.), *Introspection in second language research* (pp. 82-95). Clevedon, England: Multilingual Matters.

Cohen, A. D. (1991). Feedback on writing: The use of verbal reports. *Studies in Second Language Acquisition, 13*(2), 133-159.

Cohen, A. D. (1994a). *Assessing language ability in the classroom* (2nd ed.). Boston: Newbury House/Heinle & Heinle.

Cohen, A. D. (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: Longman.

Cohen, A. D. (2000). Exploring strategies in test taking: Fine-tuning verbal reports from respondents. In G. Ekbatani & H. Pierson (Eds.), *Learner-directed assessment in ESL* (pp. 131-145). Mahwah, NJ: Erlbaum.

Cohen, A. D. (2005, June). *Coming to terms with language learner strategies: What do strategy experts think about the terminology and where would they direct their research?* Paper presented at the 2nd International Conference of the Centre for Research in International Education, AIS St. Helens, Auckland, NZ.

Cohen, A. D., & Aphek, E. (1979). *Easifying second language learning.* Jerusalem, Israel: The Jacob Hiatt Institute. (ERIC No. ED163753).

Cohen, A. D., & Cavalcanti, M. C. (1987). Giving and getting feedback on compositions: A comparison of teacher and student verbal report. *Evaluation and Research in Education*, *1*(2), 63-73.

Cohen, A. D., & Cavalcanti, M. C. (1990). Feedback on compositions: Teacher and student verbal reports. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 155-177). Cambridge, England: Cambridge University Press.

Cohen, A. D., Glasman, H., Rosenbaum-Cohen, P. R., Ferrara, J., & Fine, J. (1979). Reading English for specialized purposes: Discourse analysis and the use of student informants. *TESOL Quarterly*, *13*(4), 551-564.

Cohen, A. D., & Hosenfeld, C. (1981). Some uses of mentalistic data in second-language research. *Language Learning, 31*(2), 285-313.

Enright, M. K., 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. K., & Schedl, M. (2000). *Reading for a reason: Using reader purpose to guide test design.* Manuscript in preparation..

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Revised ed.). Cambridge, MA: MIT Press.

ETS. (2002a). *LanguEdge Courseware: Handbook for scoring speaking and writing*. Princeton, NJ: Author.

ETS. (2002b). *LanguEdge Courseware score interpretation guide*. Princeton, NJ: Author.

ETS. (2003). *Task specifications for Next Generation TOEFL reading test.* Unpublished manuscript.

Fransson, A. (1984). Cramming or understanding? Effects of intrinsic and extrinsic motivation on approach to learning and test performance. In J. C. Alderson & A. H. Uruquhart (Eds.), *Reading in a foreign language* (pp. 86-121). London: Longman..

Grabe, W. (2004). Research on teaching reading. *Annual Review of Applied Linguistics, 24*, 44-69.

Green, A. J. F. (1998). *Using verbal protocols in language testing research: A handbook.* Cambridge, UK: Cambridge University Press.

Hosenfeld, C. (1984). Case studies of ninth grade readers. In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 231-249). London: Longman.

Kern, R. G. (1994). The role of mental translation in second language reading. *Studies in Second Language Acquisition*, *16*(4), 441-461.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363-394.

Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P., & Taylor, C. (1999). *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: Cambridge University Press.

Kirsch, I. S., & Mosenthal, P. B. (1990). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly, 25*(1), 5-30.

Laufer, B. (1991). *Similar lexical forms in interlanguage*. Tübingen, Germany: Gunter Narr.

Leow, R. P., & Morgan-Short, K. (2004). To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in Second Language Acquisition, 26*(1), 35-57.

Mosenthal, P. B., & Kirsch, I. S. (1992). Learning from exposition: Understanding knowledge acquisition from a knowledge model perspective. *Journal of Reading, 35*, 356-370.

Nevo, N. (1989). Test-taking strategies on a multiple-choice test of reading comprehension. *Language Testing*, *6*(2), 199-215.

Norusis, M. (1997). *SPSS: Guide to data analysis*. Upper Saddle River, NJ: Prentice Hall.

Porte, G. K. (2002). *Appraising research in second language learning: A practical approach to critical analysis of quantitative research.* Philadelphia: John Benjamins.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Lawrence Erlbaum.

Radford, J. (1974). Reflections on introspection. *American Psychologist, 29*(4), 245-250.

Raimes, A. (1987). Language proficiency, writing ability, and composing strategies: A study of ESL college student writers. *Language Learning*, *37*(3), 439-467.

Singhal, M. (2001). Reading proficiency, reading strategies, metacognitive awareness and L2 readers. *The Reading Matrix, 1*(1), 8.

Skehan, P. (1998). Task-based instruction. *Annual Review of Applied Linguistics*, *18*, 268-286.

Skibniewski, L. (1990). The writing processes of advanced foreign language learners: Evidence from thinking aloud and behavior protocols. In J. Fisiak (Ed.), *Papers and studies in contrastive linguistics* (pp. 193-202). Poznan, Poland: Adam Mikiewicz University.

Stemmer, B. (1991). *What's on a C-test taker's mind: Mental processes in C-test taking.* Bochum, Germany: Brockmeyer.

Swales, J. (1981). *Aspects of article introduction*. Birmingham, England: The University of Aston, Language Studies Unit.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge, England: Cambridge University Press.

Upton, T. A. (1997). First and second language use in reading comprehension strategies of Japanese ESL students. *TESL-EJ, 3*(1), 22.

Upton, T. A. (1998). "Yuk, the skin of insects!" Tracking sources of errors in second language reading comprehension. *Journal of College Reading and Learning*, *29*(1), 5-20.

Upton, T. A., & Lee-Thompson, L.-C.. (2001). The role of the first language in second language reading. *Studies in Second Language Acquisition, 23*(4), 469-495.

Uruquhart, S., & Weir, C. (1998). *Reading in a second language: Process, product and practice*. Harlow, Essex, England: Longman.

Warren, J. (1996). How students pick the right answer: A 'think aloud' study of the French CAT. In J. Burston, M. Monville-Burston, & J. Warren (Eds.), *Australian Review of Applied Linguistics* (Occasional Paper No. 15; pp. 79-94). Canberra, Australia: Applied Linguistics Association of Australia.

Williams, E., & Moran, C. (1989). Reading in a foreign language at intermediate and advanced levels with particular reference to English. *Language Teaching, 22*(4), 217-228.

Zamel, V. (1983). The composing processes of advanced ESL students: Six case studies. *TESOL Quarterly*, *17*(2), 165-187.

Notes

[1] Allan (1992) developed a test of test-wiseness for ESL students. The test included stem-option cues, in which it was possible to match information from the stem with information in the correct option; grammatical cues, where only one alternative matched the stem grammatically; similar options, where several distractors could be eliminated because they essentially said the same thing; and item giveaways, where another item already gave away the information.

[2] "Reading to find information" is not assessed in isolation in the TOEFL because the task options and level of difficulty are not considered sufficient for discriminating readers of different abilities, "unless time is constrained so that automatic processes have to be engaged . . . At present, this reading purpose is assessed in combination with reading for basic comprehension" (Enright et al., 2000, p. 31).

[3] It should be noted that other reading tasks are also included in other sections of the new TOEFL, including reading/speaking and reading/writing tasks (ETS, 2002a, 2002b). This study does not examine the reading tasks—or reading purposes—that are outside of the Reading section of the new TOEFL.

[4] In their discussion of the new TOEFL task specifications, Enright and Schedl (2000) describe other types of multiple-selection multiple-choice item types that may be considered in future versions of the test.

[5] One of the less "traditional" inferencing item types (Inferencing/Insert text) requires test takers to click on one of four squares [□] found in a given paragraph in the test passage to indicate where a new sentence would "best fit" when it is added. When a square [□] is clicked, the new sentence is automatically inserted into that context, but can be freely moved from one square [□] to another. Although the item structure and procedure are fairly innovative and reflective of the innovations computer-based testing can offer, the item is still considered a single-selection multiple-choice item because only one of four possible options is correct.

[6] It took 7–10 hours to translate/transcribe, code, double-check coding, tag, upload to computer, and burn to DVD each hour of verbal report collected.

[7] A complete list of the LanguEdge item classifications and codes, which was provided by ETS, is included as Appendix G.

[8] Consultation with Christopher Wild, the chair of the Department of Statistics at the University of Auckland, New Zealand, confirmed that there were too many variables involved to warrant a more "rigorous" statistical treatment of reading and test-taking strategy frequencies.

[9] Appendix H provides the actual frequency rates (= # occurrences/# of items of that type) for all of the strategies across all of the item types.

[10] For the purpose of the coding scheme in this study, prior familiarity with vocabulary in the passages was included as "background knowledge."

[11] For 2-point R2L items, the respondents had to get 2 out of 3 answers correct to receive one point and all 3 answers correct to receive the full 2 points. For 4 point items, the examinee had to get 4 out of 7 answers correct to receive any credit, 5 answers correct to receive 2 points, 6 answers correct to receive 3 points, and all 7 answers correct to receive the full 4 points. For answers to be correct in schematic table items, the examinee not only had to select the correct response but also place it in the correct category.

[12] We need to remind readers that this study was conducted exclusively with results based on the publicly available LanguEdge tests, which ETS would consider a prototype of the new TOEFL, rather than with the actual reading tests on the new TOEFL. Hence, the findings refer to this prototype and not to the test itself. In fact, one of the features of the new TOEFL is that it will be revised or "reversioned" as frequently as necessary, based on feedback as to how effectively it is working.

[13] The TOEFL committee reported to us that they prototyped these items with and without the text present, and found no significant performance differences. Our study most likely found the explanation for this—namely, that many of the respondents had familiarized themselves with the passage by the time that they arrived at these items. Consequently, they already had a working mental construct of the text and did not need to consult the text again at this point in the test.

List of Appendixes

Appendix A

Recruitment Flyer

Are you an International Student studying ESL?

If your answer is: Yes…

We are seeking volunteers to be a part of a research study, "Strategies in Responding to the new TOEFL Reading Tasks."

We are looking for international students who are currently enrolled in ESL programs and who fits all of the following descriptions:

1. You are a prenative speaker of English. (While we are primarily looking for students from China, Japan, and Korea, we also want to include students from other countries as well!)
2. You have an English reading proficiency that is at an intermediate, intermediate high, or advanced level.
3. You are at least 18 years old.
4. You are willing to take the reading section from the prototype of the new TOEFL and learn a new technique to describe the test-taking and reading strategies you use to complete the test.
5. You are willing to meet with a researcher for about two hours per day on three different days.

You will be compensated with a University of Minnesota bookstore gift certificate equivalent to $15/hour for your participation. So if you meet with us for six hours, you could earn a bookstore certificate worth $90 that you can spend however you want!

If you are interested or have any questions, please call or e-mail to one of us below:

[The original flyer listed four research assistants associated with the project and their contact information.]

Or, provide your name and contact information below and return to your teacher. We will then contact you.

Name: _____

E-mail: _____

Phone Number: _____

***Confidentiality will be strictly maintained.**

Appendix B

Background Questionnaire

Study of Test-Taking Strategies on the New TOEFL

Date: _____

Name: _____

Age: _____

Gender: ___ Male ___ Female

What is your native language? _____

What is your native country/culture? _____

How many years have you studied ESL? _____

Did you study ESL in your primary and/or secondary schooling?

___ Yes If yes, starting at what age? _____

___ No

How long have you been in the U.S.? _____

Are you currently in an ESL program? ___ Yes ___ No.

Have you studied ESL in the U.S. in the past? ___ Yes ___ No.

If yes, for how long? _____

How many times have you taken the TOEFL? _____

What was the month/year you last took the TOEFL? _____

Have you ever taken a TOEFL preparation class? __Yes __No

If yes, did you find it helpful? __Yes __No

If yes, why?

_____

_____

_____

_____

_____

Please self-assess your English reading ability:

___beginning ___intermediate ___high intermediate ___advanced

What is your current student status?

___ ESL only. If you were admitted to a university, what field of study would interest you?

_____

___ Undergraduate. What is your (intended) major? _____

___ Graduate. What department are you in? _____

How comfortable are you with using a computer?

___very comfortable ___comfortable ___not comfortable

Have you ever taken a reading test on a computer before?

__ No __ Yes. If yes, which test?

_____

Appendix C

Consent Form

Strategies in Responding to the New TOEFL Reading Tasks

You are invited to be in a research study of describing the strategies used in taking the Reading Section on the new TOEFL. You were selected as a possible participant because you are from one of these countries; China, Japan, or South Korea and currently studying ESL at Minnesota English Center (MEC) at the University of Minnesota or at ESL programs in Twin Cities universities with the English reading proficiency above a beginner level. We ask you that you read this form and ask any questions you may have before agreeing to be in the study.

This study is being conducted by: Professor Andrew Cohen, Department of English as a Second Language at the University of Minnesota, Associate Professor Thomas Upton, English as a Second Language Program at Indiana University-Purdue University Indianapolis, with the assistance of four research assistants; Mesut Akdere, Haesook Koo, Hitomi Maeda, and Sam Song at the College of Education and Human Development, University of Minnesota.

Background Information

The purpose of this study is to describe the strategies that respondents use on the reading section of the prototype of the new TOEFL (ETS, 2002a). Respondents will be asked to take the test using both the more traditional formats and the new selected-response items and will be asked to describe what strategies they used either in their native language or in English or both. The main goal of this study is to characterize the response strategies used to deal with the new item formats and to compare these strategies to those used on the more traditional items.

Procedures:

If you agree to be in this study, we would ask you to do the following things:

1. filling out background questionnaire that asks your previous English learning background including testing experiences of TOEFL (15 minutes)
2. participate in the orientation to the reading section of the TOEFL prototype test (20 minutes)
3. take pretest of the reading section (75 minutes)
4. doing verbal report on the last test item (up to 60 minutes)

After the four procedures above are finished and based on the information gathered, we will notify whether or not you will continue to participate in the remaining parts of this study. If you're selected to remain in this study, we will ask you to do the followings:

5. participating in the orientation to verbal training (20 minutes)
6. completing two additional reading passages from the reading section on a separate day (each 90 minutes)
7. two verbal reporting sessions after taking each reading passage (each up to 60 minutes)

For the procedure of 5, 6, and 7 above, we will video-tape all three test-taking sessions and audio-tape two verbal reporting sessions with one R.A representing your native language. There will be no preparation necessary for taking tests. Test scores will not be reported to ETS or to your ESL teachers. Therefore, your test scores from this research will not be added to your testing record in ETS regarding TOEFL.

Risks and Benefits of Being in the Study

The study has two risks: First, respondents may experience mental tiredness or anxiety due to concentration on test items during testing. To minimize this, a time limit will not be applied during each testing sessions. Second, respondents may experience physical tiredness due to taking tests on computer (computer-based testing). To minimize this, respondents will be asked to have a short break before verbal reporting.

There no direct benefits to participate in this research. However, it is envisioned that you stand to benefit from having an opportunity of practicing the reading section of the prototype of the new TOEFL (ETS, 2002a). It would appear that by practicing this prototype version of the new TOEFL, you may gain insights in preparing the new TOEFL Reading section and for your continuing study in ESL.

Compensation

You will receive payment at the end of your participation. You will be given a bookstore gift certificate for your time, which will be valued at $15/hour. You will be paid only after you complete your participation in all of the procedures mentioned above in Procedures.

Confidentiality

All data collected form participants will be handled with the strictest confidentiality. The records of this study will be kept private in a locked file and only researchers will have access to the records only using secure password in database. In any sort of report we might publish, we will not include any information that will make it possible to identify a subject. For the purposes of this research, you will be referred to by a pseudonym. Data collected and research records will be stored securely The research results will not be shared by anyone except for two main investigators, four research assistants, the sponsor of this research, Educational Testing Service (ETS), and yourself. All tape and audio recordings are accessible only by two main investigators, four research assistants, and the sponsor, ETS.

Voluntary Nature of the Study

Your decision whether or not to participate will not affect your current or future relations with the Minnesota English Center (MEC) or the University of Minnesota or your currently enrolled ESL program. If you decide to participate, you are free to not answer any question or withdraw at any time with out affecting those relationships.

Contacts and Questions

The researchers conducting this study are: Professor Andrew Cohen, Department of English as a Second Language at the University of Minnesota, Associate Professor Thomas Upton, English as a Second Language Program at Indiana University-Purdue University Indianapolis, with the assistance of four research assistants; Mesut Akdere, Haesook Koo, Hitomi Maeda, and Sam Song at the College of Education and Human Development, University of Minnesota. You may ask any questions you have now. If you have questions later, you may contact Dr. Andrew Cohen, Dr. Thomas Upton, or research assistants Mesut Akdere, Haesook Koo, Hitomi Maeda, or Sam Song. [Contact information was included in the original consent form.]

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455; (612) 624-1650.

You will be given a copy of this information to keep for your records.

Statement of Consent

I have read the above information. I have asked questions and have received answers. I consent to participate in the study.

Signature:_____ Date: _____

Signature of Investigator:_____ Date: _____

Appendix D

Verbal Report Training

NOTE: [Items in bold and in brackets are for the RAs and should be deleted from the form given to students.]

[This should be translated in Chinese/Korean/Japanese for each subgroup]

[The RA should read these directions aloud while the subject follows along with the written copy.]

When you take a reading test, your mind is very active. That's why we often get tired when we take tests; it's hard work. You have to understand what you are reading, and you have to try to figure out how to answer the multiple-choice questions. But you usually don't talk during tests, so no one else knows what you are thinking or doing as you try to answer a question.

In this study, I want to know what you are thinking and doing as you work your way through each item of the test. I am going to ask you to <u>think out loud</u> and describe everything that you are thinking about and doing as you deal with each multiple-choice item. In other words, we want to know your reasoning for selecting the option that you consider to be the correct or best answer, and for eliminating the other options.

This may feel a little strange at first, but it might help to imagine that you are in a room by yourself. With a little practice, you will become more comfortable talking out loud about what you are thinking.

It is important that you talk as much as possible. We are doing this study to better understand what you are thinking about and doing as you do each item. We can only know what you are thinking about and doing if you talk out loud as you do it. If you are silent for more than a few seconds, I will say, "What are you thinking" in order to remind you to talk. Of course, your thoughts are going to be in both (Chinese/Japanese/Korean) and English. You should think out loud in whatever language you are thinking. If you are thinking in (Chinese/Japanese/Korean), then you should speak in (Chinese/Japanese/Korean). If you are thinking in English, then you should speak in English. We want to know not only what you are doing, but why you are doing it. As you think out loud, I may ask you to explain what you are thinking further if it is not clear why you are doing what you are doing. For example, as you are looking at a test question you

say out loud, “Letter A is not the right answer,” I will ask you, “Why did you decide that?” After a little practice, you should understand what it is we are asking you to do.

I will model for you an example of someone thinking out load as they read and try to answer a question on a reading test.

[The test taker is looking at this question and text]

> The railroad was not the first institution to impose regularity on society, or to draw attention to the importance of precise timekeeping. For as long as merchants have set out their wares at daybreak and communal festivities have been celebrated, people have been in rough agreement with their neighbors as to the time of day. The value of this tradition is today more apparent than ever. Were it not for public acceptance of a single yardstick of time, social life would be unbearably chaotic: the massive daily transfers of goods, services, and information would proceed in fits and starts; the very fabric of modern society would begin to unravel.
>
> What is the main idea of the passage?
>
> A. In modern society we must make more time for our neighbors.
> B. The traditions of society are timeless.
> C. An accepted way of measuring time is essential for the smooth functioning of society.
> D. Society judges people by the times at which they conduct certain activities.

[The underlined portion should be said in the L1 of the respondents]

OK, I’m going to look at the question first before I read. What is main idea? OK, I’ll read them. “A.” more time for neighbors, more time for neighbors. “B.” traditions, timeless. OK, I’ll go to “C” smooth functioning of society. I’m not sure what that means. Read on. “D.” by the times at which they conduct certain activities. That makes no sense. OK, I’ll read the passage.

First sentence. impose regularity, I don’t know what that means but I think it’s a noun because of how it’s used; I’ll keep reading. Importance of precise timekeeping. OK, that means to be accurate, on time. I’ll go on. Wares, what’s that? I’ll read again to see if I can understand it better. OK, not important, just says people have common idea of time. Keep going. This is valuable? Single yardstick of time…huh? Again to better understand it, ok we need common time or something happens.

OK what does it say, this has something to do with the importance of keeping track of time. Look at questions. "A" more time for neighbors. No, the text is not about making more time for neighbors. "B" says traditions… don't think so. It is not about traditions in general. "C" accepted way of measuring time…It looks pretty good. Let me come back. "D" society judges people… That sounds pretty silly. The text isn't really about society judging people by the times in which they conduct certain activities. So, now let's look at the choice "C" again. In the passage it says, "Public acceptance of a single yardstick of time" and this is just another way of saying "an accepted way of measuring time" which is what it says in choice "C." But is it the main idea? Yeah, it appears to be general enough. I think it's right, so I'll choose it.

OK. Here's another example of an item using this same passage, and my verbal report about it:

In line 5, the phrase "this tradition" refers to

A. the practice of starting the business day at dawn
B. friendly relations between neighbors
C. the railroad's reliance on time schedules
D. people's agreement on the measurement of time

Let's see. What does "this tradition" refer to? This passage is about the importance of time. "A" is too limited, I think. We're not just talking about starting the business day. No. The same with "B" about neighbors and "C" about railroads. I'll go with "D" since it refers to the measurement of time altogether. The other three choices are just examples of the use of time.

Do you understand what I am asking you to do? Do you have any questions?

Let's do a practice question.

Remember,

- you want to say out loud everything that you are thinking about as you read and answer the question.
- if you are thinking in English, then speak in English. If you are speaking in (Chinese/Japanese/Korean), then speak in (Chinese/Japanese/Korean).
- if you are quiet for more than a few seconds, I will ask you, "What are you thinking?"

- I will ask you to describe your thoughts more if I want to get more information on what you are thinking.
- I will not be able to help you at all with the questions.
- I will be testing the recording equipment while you practice so that you can get used to that, too.

Please read the following passage and answer the question after it.

Test Item:

(Start thinking aloud now)

[The test item should be left in English]

[Be sure to ask students "What are you thinking?" whenever they fall silent for more than a few second.]

Passage: "…These laws are universal in their application, regardless of cultural beliefs, geography, or climate. If pots have no bottoms or have large openings in their sides, they could hardly be considered containers 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. . ."

The word they in the passage refers to

A. applied-art objects
B. the laws of physics
C. containers
D. the sides of pots

Do you feel comfortable thinking aloud? You can practice on another question in you would like.

[RAs should have subjects do second example to get more practice if they don't feel the subject quite has the verbal report process down.]

(Starting thinking aloud now)

[The test item should be left in English]

[Be sure to ask students "What are you thinking?" whenever they fall silent for more than a few second.]

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 in 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

Do you have any questions about what we want you to do?
Are you ready to begin?

[Go to Testing Protocol]

Appendix E

Item Type Descriptions

NOTE: Descriptions and examples in this section are all taken from ETS, 2003, unless otherwise noted. In the examples, the correct answer is noted with an "*."

Basic Comprehension

These items focus on "an examinee's ability to understand important information in a text based on the lexical, syntactic, and semantic content of the text" (p. 4).

Type 1: Vocabulary Items

"These items measure examinee's ability to comprehend the meanings of individual words and phrases as used in the context of the passage" (p. 4).

EXAMPLE (p. 4):

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

Type 2: Reference Items

"These items measure examinee's ability to identify relationships between pronouns and other anaphoric devices and their antecedents/postcedents within the passage" (p. 6).

EXAMPLE (p. 7):

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

Type 3: Sentence Simplification Items

“These items measure 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 (p. 9):

[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 . . ."

Type 4: Factual Information Items

"These items measure 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 (p. 11):

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

Type 5: Negative Fact Items (also called Not/Except Items)

"These items 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. 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 (p. 12):

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

Inferencing

“Inferencing tasks share some characteristics with both basic comprehension tasks and reading to learn tasks. While they can still be used to test sentence-level information, as basic comprehension items do, they can also be used to test information across multiple parts of the text. They may also require abilities related to connecting information and recognizing the organization and purpose of the text” (p. 25).

Type 1: Inferencing Items

“These items measure 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 (p. 26):

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

Type 2: Rhetorical Purpose Items

"These items 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 or exposition and the author's rhetorical purpose" (p. 27).

EXAMPLE (p. 29):

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

Type 3: Insert Text Items

"These items measure 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 (p. 33):

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.

Reading to Learn

"Reading to learn is seen as involving more than understanding discrete points and getting the general idea based on the lexical, syntactic, and semantic content of texts. It also involves

- recognizing the organization and purpose of the text
- conceptualizing and organizing text information into a mental framework
- distinguishing major from minor ideas and essential from nonessential information

- understanding rhetorical functions such as cause-effect relationships, compare-contrast relationships, arguments, etc.

The questions require readers to show evidence of being able to both comprehend individual points and use a framework such as a schematic table or a summary table to reconstruct the major ideas and important supporting information from the text. Having an organized mental representation of the text is seen as critical to learning from the text because it allows the reader to remember important information from the text and apply it in new situations" (p. 14).

Type 1: Prose Summary Items

"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 (p. 17):

[***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.

Type 2: Schematic Table Items

"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 F

Strategy Codes

1. Subject ID#

Chinese Subjects

C1, C2, C3, C4, C5, C6, C7, C8

Japanese Subjects

J1, J2, J3, J4, J5, J6, J7, J8

Korean Subjects

K1, K2, K3, K4, K5, K6, K7, K8

Other Subjects

O1, O2, O3, O4, O5, O6, O7, O8

2. Item

Test 1, Passage 1, Questions 1-13

T1P1Q1, T1P1Q2, T1P1Q3, T1P1Q4, T1P1Q5, T1P1Q6, T1P1Q7, T1P1Q8, T1P1Q9, T1P1Q10, T1P1Q11, T1P1Q12, T1P1Q13

Test 1, Passage 2, Questions 1-13

T1P2Q1, T1P2Q2, T1P2Q3, T1P2Q4, T1P2Q5, T1P2Q6, T1P2Q7, T1P2Q8, T1P2Q9, T1P2Q10, T1P2Q11, T1P2Q12, T1P2Q13

Test 1, Passage 3, Questions 1-12

T1P3Q1, T1P3Q2, T1P3Q3, T1P3Q4, T1P3Q5, T1P3Q6, T1P3Q7, T1P3Q8, T1P3Q9, T1P3Q10, T1P3Q11, T1P3Q12

Test 2, Passage 1, Questions 1-13

T2P1Q1, T2P1Q2, T2P1Q3, T2P1Q4, T2P1Q5, T2P1Q6, T2P1Q7, T2P1Q8, T2P1Q9, T2P1Q10, T2P1Q11, T2P1Q12, T2P1Q13

Test 2, Passage 2, Questions 1-13

T2P2Q1, T2P2Q2, T2P2Q3, T2P2Q4, T2P2Q5, T2P2Q6, T2P2Q7, T2P2Q8, T2P2Q9, T2P2Q10, T2P2Q11, T2P2Q12

Test 2, Passage 3, Questions 1-13

T2P3Q1, T2P3Q2, T2P3Q3, T2P3Q4, T2P3Q5, T2P3Q6, T2P3Q7, T2P3Q8, T2P3Q9, T2P3Q10, T2P3Q11, T2P3Q12, T2P3Q13

3. Question Type

Basic Comprehension

BC-v (=vocabulary)

BC-f (=fact)

BC-n/e (=not/except)

BC-pr (=pronoun reference

BC-ss (=sentence simplification)

Inferencing

I (=basic inference)

I-it (=insert text)

I-rp (=rhetorical purpose)

Reading-to-Learn

R2L-ps (=prose summary)

R2L-st (=schematic table

4. Strategy Type

Reading

R1 = plans a goal for the passage

R2-R22 = (Use Reading Strategies Coding Rubric—Figure 1)

Test-Management Strategies

T1 = goes back to the question and rereads it

T2-T28 = (Use Test-Taking Strategies Coding Rubric—Figure 2)

Test-Wiseness Strategies

TW1-TW3 = (Use Test-Taking Strategies Coding Rubric—Figure 3)

5. Language of Verbal Report

L1 = first language (also in **bold**)

6. Answer Correct

AC = answer correct

AW = answer wrong

PC = answer partially correct (for R2L questions: 2/3 correct on R2L-ps; 3/4 or 4/5 correct on R2L-st)

Example:

[K1]

<T1P2Q12>

<I-it>

<T4> Square . . . "the new party…the economy" [He reads the sentence to be inserted] </T4>

[He scrolls down the text to adjust the location of the second paragraph]

<R6> "During the Jackson's second term . . ." "The Democrats tented to view . . ." "This paper . . ." </R6>

[He goes to the beginning of this paragraph]

<T5b>"During the Jackson's second term . . . the Whig party" </T5b> </L2>

<T11d> I think the first square is the answer. <L1> Because from the second square, they all talk about the Democrats or are not related to the sentence needed to be inserted. </L1> </T11d>

<T9> Well, wait a second, right. </T9>

<T11b> <L1> I think the first square is the answer. </L1> </T11b>

</I-it>

</T1P2Q12>

[AC]

Appendix G

Desciption of LanguEdge Test Items

Displayed item #	Set	Item # w/in item type	Item type	Item code	Accession #	Content code	Classification
Test 2							
39	1	1	Reading	Desert1	VB442904	Vocabulary	Basic comprehension
40	1	2	Reading	Desert2	VB442908	Fact	Basic comprehension
41	1	3	Reading	Desert3	VB442912	Vocabulary	Basic comprehension
42	1	4	Reading	Desert4	VB444363	Fact	Basic comprehension
43	1	5	Reading	Desert5	VB444502	Vocabulary	Basic comprehension
44	1	6	Reading	Desert6	VB444504	Fact	Basic comprehension
45	1	7	Reading	Desert7	VB442916	Vocabulary	Basic comprehension
46	1	8	Reading	Desert8	VB444222	Fact	Basic comprehension
47	1	9	Reading	Desert9	VB444506	Not/except	Basic comprehension
48	1	10	Reading	Desert10	VB445527	Sentence simplification	Basic comprehension
49	1	11	Reading	Desert11	VB444516	Inference	Inference
50	1	12	Reading	Desert12	VB444519	Insert text	Inference
51	1	13	Reading	Desert13	VB444680	Prose summary	Reading to learn
52	1		Reading/ speaking		VB444963	Constructed Response	

(Table continues)

Table (continued)

Displayed item #	Set	Item # w/in item type	Item type	Item code	Accession #	Content code	Classification
53	2	1	Reading	Cinema1	VB442845	Not/except	Basic comprehension
54	2	2	Reading	Cinema2	VB442850	Rhetorical purpose	Inference
55	2	3	Reading	Cinema3	VB445529	Sentence simplification	Basic comprehension
56	2	4	Reading	Cinema4	VB442905	Vocabulary	Basic comprehension
57	2	5	Reading	Cinema5	VB442906	Vocabulary	Basic comprehension
58	2	6	Reading	Cinema6	VB442907	Fact	Basic comprehension
59	2	7	Reading	Cinema7	VB442917	Fact	Basic comprehension
60	2	8	Reading	Cinema8	VB442921	Fact	Basic comprehension
61	2	9	Reading	Cinema9	VB444224	Pronoun reference	Basic comprehension
62	2	10	Reading	Cinema10	VB444226	Fact	Basic comprehension
63	2	11	Reading	Cinema11	VB444227	Vocabulary	Basic comprehension
64	2	12	Reading	Cinema12	VB444229	Insert text	Inference
65	2	13	Reading	Cinema13	VB444366	Prose summary	Reading to learn
66	2		Reading/ writing		VB444380	Constructed response	

(Table continues)

Table (continued)

Displayed item #	Set	Item # w/in item type	Item type	Item code	Accession #	Content code	Classification
67	3	1	Reading	Cetacean1	VB442846	Fact	Basic comprehension
68	3	2	Reading	Cetacean2	VB442903	Inference	Inference
69	3	3	Reading	Cetacean3	VB442910	Vocabulary	Basic comprehension
70	3	4	Reading	Cetacean4	VB442919	Fact	Basic comprehension
71	3	5	Reading	Cetacean5	VB442920	Pronoun reference	Basic comprehension
72	3	6	Reading	Cetacean6	VB444230	Vocabulary	Basic comprehension
73	3	7	Reading	Cetacean7	VB444235	Fact	Basic comprehension
74	3	8	Reading	Cetacean8	VB444238	Inference	Inference
75	3	9	Reading	Cetacean9	VB444240	Inference	Inference
76	3	10	Reading	Cetacean10	VB445528	Sentence simplification	Basic comprehension
77	3	11	Reading	Cetacean11	VB444365	Vocabulary	Basic comprehension
78	3	12	Reading	Cetacean12	VB444367	Insert text	Inference
79	3	13	Reading	Cetacean13	VB444371	Prose summary	Basic comprehension [*sic*]

(Table continues)

Table (continued)

Displayed item #	Set	Item # w/in item type	Item type	Item code	Accession #	Content code	Classification
Test 1							
40	1	1	Reading	express1	VB441660	Vocabulary	Basic comprehension
41	1	2	Reading	express2	VB441661	Rhetorical purpose	Inference
42	1	3	Reading	express3	VB442842	Vocabulary	Basic comprehension
43	1	4	Reading	express4	VB442843	Pronoun reference	Basic comprehension
44	1	5	Reading	express5	VB442909	Fact	Basic comprehension
45	1	6	Reading	express6	VB445532	Sentence simplification	Basic comprehension
46	1	7	Reading	express7	VB442855	Fact	Basic comprehension
47	1	8	Reading	express8	VB442857	Fact	Basic comprehension
48	1	9	Reading	express9	VB442913	Vocabulary	Basic comprehension
49	1	10	Reading	express10	VB442914	Vocabulary	Basic comprehension
50	1	11	Reading	express11	VB442918	Fact	Basic comprehension
51	1	12	Reading	express12	VB444223	Insert text	Inference
52	1	13	Reading	express13	VB444225	Prose summary	Reading to learn
53	1		Reading/ Speaking	express14	VB444369	Constructed response	

(Table continues)

Table (continued)

Displayed item #	Set	Item # w/in item type	Item type	Item code	Accession #	Content code	Classification
54	2	1	Reading	politics1	VB444231	Vocabulary	Basic comprehension
55	2	2	Reading	politics2	VB444234	Fact	Basic comprehension
56	2	3	Reading	politics3	VB444241	Fact	Basic comprehension
57	2	4	Reading	politics4	VB444962	Fact	Basic comprehension
58	2	5	Reading	politics5	VB444964	Fact	Basic comprehension
59	2	6	Reading	politics6	VB444965	Vocabulary	Basic comprehension
60	2	7	Reading	politics7	VB444966	Fact	Basic comprehension
61	2	8	Reading	politics8	VB444967	Vocabulary	Basic comprehension
62	2	9	Reading	politics9	VB444968	Inference	Inference
63	2	10	Reading	politics10	VB444969	Not/except	Basic comprehension
64	2	11	Reading	politics11	VB445531	Sentence simplification	Basic comprehension
65	2	12	Reading	politics12	VB445011	Insert text	Inference
66	2	13	Reading	politics13	VB445012	Prose summary	Reading to learn
67	2		Reading/ Writing	politics14	VB445013	Constructed response	

(Table continues)

Table (continued)

Displayed item #	Set	Item # w/in item type	Item type	Item code	Accession #	Content code	Classification
68	3	1	Reading	geo1	VB441659	Fact	Basic comprehension
69	3	2	Reading	geo2	VB442847	Vocabulary	Basic comprehension
70	3	3	Reading	geo3	VB442844	Inference	Inference
71	3	4	Reading	geo4	VB442848	Vocabulary	Basic comprehension
72	3	5	Reading	geo5	VB442849	Fact	Basic comprehension
73	3	6	Reading	geo6	VB442851	Rhetorical purpose	Inference
74	3	7	Reading	geo7	VB442853	Vocabulary	Basic comprehension
75	3	8	Reading	geo8	VB442854	Pronoun reference	Basic comprehension
76	3	9	Reading	geo9	VB445530	Sentence simplification	Basic comprehension
77	3	10	Reading	geo10	VB442859	Fact	Basic comprehension
78	3	11	Reading	geo11	VB442860	Insert text	Inference
79	3	12	Reading	geo12	VB442861	Table	Reading to learn

Note. Data provided by ETS.

Appendix H

Frequency of Reported Use of Reading and Test-Taking Strategies

Strategy	BC-v	BC-pr	BC-ss	BC-f	BC-n/e	I	I-rp	I-it	R2L-ps	R2L-st
R1	0.00	0.03	0.00	0.09	0.00	0.07	0.04	0.04	0.00	0.00
R2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00
R3	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
R4	0.00	0.00	0.03	0.11	0.00	0.00	0.08	0.00	0.00	0.00
R5	0.02	0.00	0.03	0.02	0.00	0.00	0.00	0.02	0.00	0.00
R6	0.67	1.67	1.54	1.83	2.23	1.82	1.66	1.51	0.70	1.00
R7	0.17	0.48	0.19	0.38	0.23	0.18	0.30	0.67	0.65	0.70
R8	0.00	0.00	0.00	0.03	0.00	0.00	0.04	0.07	0.13	0.00
R9	0.48	0.78	0.73	1.23	0.58	0.93	0.85	0.51	0.33	1.20
R10	0.08	0.03	0.22	0.45	0.15	0.07	0.23	0.04	0.06	0.00
R11	0.03	0.00	0.05	0.06	0.00	0.09	0.00	0.00	0.02	0.00
R12	0.02	0.06	0.00	0.00	0.00	0.02	0.00	0.04	0.00	0.00
R13	0.02	0.06	0.05	0.02	0.08	0.02	0.00	0.06	0.02	0.00
R14	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R15	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00
R16	0.00	0.03	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00
R17	0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R18	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R19	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.10
R20	0.00	0.00	0.00	0.02	0.00	0.02	0.00	0.00	0.02	0.00
R21	0.00	0.00	0.11	0.00	0.00	0.00	0.04	0.02	0.00	0.00
R22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00
R23	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R24	0.00	0.00	0.00	0.02	0.08	0.00	0.00	0.00	0.00	0.00
R25	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00

(Table continues)

Table (continued)

Strategy	BC-v	BC-pr	BC-ss	BC-f	BC-n/e	I	I-rp	I-it	R2L-ps	R2L-st
R26	0.00	0.54	0.00	0.00	0.00	0.02	0.00	0.09	0.04	0.00
R27	0.02	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00
R28	0.02	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
T1	0.33	0.31	0.41	0.81	0.54	0.69	0.58	0.36	0.35	0.50
T2	0.25	0.09	0.30	0.55	0.31	0.51	0.31	0.35	0.33	1.30
T3	0.06	0.00	0.11	0.05	0.00	0.42	0.08	0.02	0.04	0.10
T4	0.13	0.00	0.08	0.20	0.27	0.24	0.19	0.07	0.85	0.90
T5	1.50	2.44	2.16	2.89	2.50	2.38	2.27	1.60	0.02	0.05
T6	0.23	0.19	0.16	0.19	0.08	0.22	0.31	0.06	0.07	0.10
T7	0.19	0.06	0.00	0.08	0.12	0.09	0.08	0.06	0.09	0.00
T8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.49	0.02	0.00
T9	0.06	0.00	0.00	0.05	0.19	0.02	0.19	0.04	0.15	0.10
T10	0.69	0.38	0.00	0.02	0.00	0.00	0.00	0.04	0.00	0.10
T11	0.00	0.03	0.00	0.00	0.00	0.02	0.04	0.02	0.02	0.10
T12	0.52	0.25	0.84	0.50	0.81	0.73	0.58	0.26	0.15	0.10
T13	0.50	0.00	0.00	0.00	0.04	0.02	0.00	0.00	0.07	0.00
T14	0.27	0.00	0.35	0.39	0.46	0.58	0.46	0.16	0.57	0.00
T15	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.42	0.02	0.00
T16	0.75	0.59	1.03	1.45	1.35	1.13	2.19	0.09	2.93	1.90
T17	0.05	0.06	0.46	0.17	0.19	0.07	0.38	0.06	0.37	0.50
T18	0.17	0.22	0.08	0.09	0.04	0.16	0.08	0.06	0.11	0.00
T19	0.11	0.09	0.35	0.13	0.31	0.24	0.12	0.29	0.52	0.00
T20	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00
T21	0.50	0.06	0.03	0.05	0.04	0.11	0.00	0.09	0.02	0.10
T22	0.45	1.13	0.95	0.89	0.88	0.93	0.96	0.96	3.30	2.80
T23	0.02	0.00	0.00	0.02	0.04	0.07	0.00	0.00	0.00	0.00
T24	0.22	0.16	0.35	0.34	0.46	0.38	0.15	0.15	0.11	0.00
T25	0.00	0.00	0.05	0.03	0.00	0.00	0.00	0.04	0.02	0.00

(Table continues)

Table (continued)

Strategy	BC-v	BC-pr	BC-ss	BC-f	BC-n/e	I	I-rp	I-it	R2L-ps	R2L-st
T26	0.00	0.13	0.03	0.00	0.00	0.07	0.04	0.31	0.00	0.00
T27	0.39	0.00	0.05	0.03	0.04	0.07	0.00	0.11	0.11	0.00
T28	1.03	1.07	2.68	1.86	1.92	2.24	2.08	0.84	2.67	1.00
TW1	0.09	0.00	0.05	0.00	0.08	0.00	0.04	0.00	0.00	0.00
TW2	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.02	0.00
TW3	0.02	0.00	0.00	0.02	0.00	0.02	0.00	0.00	0.06	0.00

Note. Strategies are ranked and grouped according to three levels of frequency rate (number of occurrences / number of items of that type): very high (≥ 1.0), high (≥ .50), and moderate (≥ .03).

Test of English as a Foreign Language
PO Box 6155
Princeton, NJ 08541-6155
USA

To obtain more information about TOEFL programs and services, use one of the following:

Phone: 1-877-863-3546
(US, US Territories*, and Canada)

1-609-771-7100
(all other locations)

Email: toefl@ets.org

Web site: www.ets.org/toefl

* America Samoa, Guam, Puerto Rico, and US Virgin Islands