‘I want to go back to the text’: Response strategies on the reading subtest of the new TOEFL®[1]

Andrew D. Cohen and Thomas A. Upton

This study describes the reading and test-taking strategies that test takers used on the ‘Reading’ section of the LanguEdge *Courseware* (2002) materials 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 skill 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, and subjects’ verbal reports accompanying items representing each of the ten item types were evaluated to determine strategy use. The findings provide insights into the response behaviors prompted by the reading tasks on the *new TOEFL*.

I BACKGROUND

1. ESL reading comprehension

In the TOEFL Monograph *TOEFL 2000 Reading Framework: A Working Paper* Enright, Grabe, Koda, Mosenthal, Mulcahy-Ernt, and Schedl (2000) outlined 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*, considered 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. Enright et al.(2000) noted other variables that can influence how first-language (LI) readers go about trying to understand an academic text and how successful those efforts will be: cognitive processing abilities (e.g., working memory efficiencies), text type (e.g., expository vs. narrative), reading task, strategy use, affect (e.g., motivation,

[1] At the time of this study, the test was referred to as the “new TOEFL”, but more recently it has come to be termed the Internet-based TOEFL or the TOEFL iBT

anxiety), topic, and L1, among others. The interplay among these variables influences how individual respondents per-form on given reading tasks as they seek to achieve a particular goal or purpose.

2 ESL reading and test-taking strategies

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 *et al.*, 1991: 42). In short, as Cohen (1994a) has noted, '[i]n 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) added, '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.

a Reading strategies: 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' (Carrell and Grabe, 2002: 234). While much of the reading process is 'automatic' in nature–which is defined as reading 'skill' (Williams and Moran, 1989: 223) – and so is 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 and Moran, 1989: 98; Urquhart and 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 (Cohen, 2005).

In keeping with our understanding of reading as a problem-solving process, reading strategy analysis provides us insights as to how readers interact with the text and how their choice of strategies influences their comprehension of the text. In their book, *Verbal protocols of reading: The nature of constructively responsive reading*, Pressley and Afflerbach (1995) grouped reading strategies into three broad categories: (1) planning and identifying strategies which help in constructing the meaning of the text; (2) monitoring strategies which serve to regulate comprehension and learning; and (3) 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 and Lee-Thompson, 2001; Carrell and Grabe, 2002).

b Test-taking strategies: Test-taking strategies are defined as those test-taking processes which the respondents have selected and which they are conscious of, at least to some degree. As noted above, the notion of strategy implies an element of selection. 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 short-cuts 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 to circumvent the need to tap their actual language knowledge or lack of it, consistent with Fransson's (1984) assertion that respondents may not proceed via the text but rather around it. In the majority of testing situations, how-ever, test-taking strategies do not lead to opting out or to the use of short cuts. In any event, as long as the language task is part of a test, students may find themselves using strategies that they would not use under non-test conditions. It is for this reason that during the pilot phase of test development it is crucial for test constructors to find out what their tests are

actually measuring.

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 and 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 (1998), '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).

3 Testing academic reading comprehension

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 *et al.*, 1999: 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. The *new TOEFL* reading task specifications (ETS, 2003) put the focus on three broad categories of reading skills: basic comprehension, reading to learn, and inferencing – each with multiple types – which form the basis for the reading portion of the *new TOEFL*.[2]

The *new TOEFL* draws on ten item types representing five Basic Comprehension tasks, three Inferencing tasks, and two Reading to Learn tasks.

In addition, attention was given to text length and text type in the reading section of the new TOEFL, which 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). 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: (1) exposition, (2) argumentation, and (3) historical biographical/autobiographical narrative. Each of these has at least one or more major text structures, such as classification, comparison/contrast, cause/effect, and problem/solution, with information presented from more than one perspective or point of view (ETS, 2003).

LanguEdge Courseware, introduced in 2002 to acquaint test users with the new TOEFL tasks, includes two prototype test forms. The reading sections of these prototype test forms include both traditional multiple-choice items as well as examples of novel multiple-selection multiple-choice items – prose summaries and schematic tables.[3] For the prose summary, test takers are asked to '[c]omplete a summary of a text, one or two sentences

[2] 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 (LanguEdge Courseware, 2002). This study does not examine the reading tasks – or reading purposes – that are outside of the 'Reading' section of the *new TOEFL*.

[3] 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.

of which are provided', by selecting three additional sentences from a list of six that express the most important ideas in the passage (Enright and Schedl, 2000: 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 and Schedl, 2000: 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.

Most importantly, these new formats were expected to elicit somewhat different 'academic-like approaches' to reading than those elicited by the more traditional formats. It was predicted that in order to respond successfully to these innovative formats, respondents would need strategies for perceiving the overall meaning of these lengthier passage, which in turn would call for 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.

4 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: 70). The purpose of this study was 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 (2002) materials developed to introduce the design of the new TOEFL. This study sought 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. Our guiding research question was: 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 10 different test item types? *Andrew D. Cohen and Thomas A. Upton* 215

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?

II. METHODOLOGY

1 Sample

Thirty-two high-intermediate to advanced non-native speakers of English representing four language groups (Chinese, Japanese, Korean, and 'Other') were recruited to participate in the study. 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 well as their reading score on the reading section of the pretest version of the LanguEdge Courseware materials. The mean score for the 32 participants on the timed LanguEdge reading pre-test was 18.9 (out of 25), which places them at about the 75th percentile in relation to ETS' 2002 Field Study (LanguEdge Courseware Score Interpretation Guide, 2002). The mean scores for the participants 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).

2 Instrumentation

As noted above, the LanguEdge Courseware materials, which repre-sent the format of the *new TOEFL* reading section, were used in this study. These reading tests use 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. These item types include the following (see Appendix A for a brief description of each, and one example of a Reading to Learn-prose summary item):

Basic Comprehension

- Basic comprehension – vocabulary (BC-v)
- Basic comprehension – pronoun reference (BC-pr)
- Basic comprehension – sentence simplification (BC-ss)
- Basic comprehension – factual information (BC-f)
- Basic comprehension – negative fact (BC-n/e)

Table 1 Test-taker characteristics

ID No.	L1	Sex	Age	Education status	Discipline	LOR† (mths)	Pretest score*
Chinese							
C1	Chinese	M	30	Graduate	Biostatistics	36	17
C2	Chinese	F	25	Graduate	Statistics	24	21
C3	Chinese	F	37	Graduate	Educational Policy/Admin	6	21
C4	Chinese	M	28	Graduate	Computer Science	60	15
C5	Chinese	M	24	Graduate	Mathematics	20	17
C6	Chinese	M	29	Graduate	Electrical Engineering	48	18
C7	Chinese	F	25	Graduate	Statistics	36	17
C8	Chinese	F	24	Undergraduate	Economics	10	19
Japanese							
J1	Japanese	F	21	Undergraduate	Linguistics	5	20
J2	Japanese	F	22	Undergraduate	English	6	15
J3	Japanese	F	23	Undergraduate	Civil Engineering	4	15
J4	Japanese	M	20	Undergraduate	Mechanical Engineering	4	22
J5	Japanese	M	34	Graduate	Business	19	20
J6	Japanese	F	22	Graduate	Educational Psycholoy	8	21
J7	Japanese	F	33	Graduate	Public Policy	29	20
J8	Japanese	F	24	Graduate	(Not Given)	19	16
Korean							
K1	Korean	M	28	Graduate	Economics	5	17
K2	Korean	F	22	Undergraduate	Global Studies	48	20
K3	Korean	F	20	Undergraduate	Biochemistry	48	20
K4	Korean	M	28	Undergraduate	Economics	60	11
K5	Korean	M	27	ESL-only	ESL	9	15
K6	Korean	F	30	Graduate	Education	48	20
K7	Korean	M	20	Undergraduate	Marketing	36	18
K8	Korean	F	25	Graduate	Computer Science	18	18

Other							
O1	Turkish	M	26	Graduate	Civil Engineering	48	22
O2	Turkish	M	26	Graduate	Mathematics	48	23
O3	Thai	F	23	Graduate	Applied Linguistics/TESL	24	20
O4	Bengali	M	27	Graduate	Finance	84	23
O5	Arabic	M	22	Undergraduate	Mechanical Engineering	40	20
O6	Turkish	M	25	Graduate	Mathematics	36	22
O7	Arabic	M	24	Undergraduate	Undeclared	36	21
O8	Turkish	M	26	Graduate	Curriculum & Instruction	24	21

†LOR = length of residence in the USA.
*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).

Inferencing

Inference – inferencing (I)

Inference – rhetorical purpose (I-rp)

Inference – insert text (I-it)

Reading to Learn

Reading to Learn – prose summary (R2L-ps)

Reading to Learn – schematic table (R2L-st)

While all the reading tests in this study contained all three general categories of item types, the distribution of specific item types varied from test to test. Nevertheless, each had at least eight Basic Comprehension items and two Inferencing items, but no more than one Reading to Learn item.

3 Data collection procedures

Participants were assigned one or the other form of the LanguEdge version of the *new TOEFL*, under regular time constraints, as a pretest to determine general reading proficiency. 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 with 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.

Since there are two forms of the LanguEdge tests, participants received in a counterbalanced fashion one as the pretest and the other as the form for which a protocol was obtained. Each pretest included three sets of readings, each one consisting of a ~600-word text with 12-13 test items accompanying it. Participants had 25 minutes to complete each of the three sets. In the first of two sessions, participants were trained in how to give concurrent verbal report (think-aloud protocol). In the second session, they completed two of three sets of texts and test items from the alternate LanguEdge version, verbalizing their reading and test-taking strategies (both audio- and digital video-recording of verbal reports), with no time constraint in order to facilitate the collection of verbal report data. Guidelines developed by Cohen (1984, 1991, 2000), Ericsson and Simon (1993), and Green (1998) for eliciting introspective, mentalistic think-alouds were used. 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.

4 Data analysis

a 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. 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. Seven Basic Comprehension, four Inferencing, and two Reading to Learn questions were selected for analysis per participant, based on a set of predetermined criteria.[4]

[4] Distribution of the items to be analyzed was determined by considering a variety of factors, including the following: (a) each general item type (basic comprehension, inferencing, reading-to-learn) should be represented in both of the reading test sets completed by each subject; (b) when possible, each of the ten item sub-types (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; (c) all questions should be used at approximately the same rate as the other questions of the same item type across all six reading test sets; and (d) verbal reports should be provided on all questions for all language groups (Chinese, Japanese, Korean, Other) across all six reading test sets.

b Strategy coding: Drawing on the literature reviewed above on reading strategies and test-taking strategies, rubrics for reading and test-taking strategies (test-management and test-wiseness) were developed to code the verbal reports. These codes were modified after the pilot study to better reflect the strategies actually used by the respondents. The rubrics used for the reading, test-management, and test-wiseness strategies in the analysis of the verbal reports can be found in Tables 2, 3, and 4.

Table 2 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 pre-reading.
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.

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 3 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 unknown vocabulary.
T10	Considers the options and checks the vocabulary option in context.
T11	Considers the options and focuses on a familiar option.
T12	Considers the options and selects preliminary option(s) (lack of certainty indicated).
T13	Considers the options and defines the vocabulary option.
T14	Considers the options and paraphrases the meaning.
T15	Considers the options and drags and considers the new sentence in context (I-it).
T16	Considers the options and postpones consideration of the option.
T17	Considers the options and wrestles with the option meaning.
T18	Makes an educated guess (e.g., using background knowledge or extra-textual knowledge).
T19	Reconsiders or double checks the response.
T20	Looks at the vocabulary item and locates the item in context.
T21	Selects options through background knowledge.
T22	Selects options through vocabulary, sentence, paragraph, or passage overall meaning (depending on 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) 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 4 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.

c Procedures for quantitative analysis: While this study was 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, 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 clearly indicated by the subjects' actions (as videotaped, e.g., specifically returning to the text to look for the answer) was tallied. Consequently, a single test question sometimes prompted 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, and analyzed, the challenge was to devise a system for rigorously distinguishing levels of frequency of occurrence. 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.[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 some quantitative measure of intervals. The cut-off points used were as follows:

very high (VH) frequency Ž1.00
high (H) frequency Ž0.50
moderate (M) frequency Ž0.30
low (L) frequency ₃ 0.29

[5] For example, strategy T5 was used a total of 107 times across the 45 different occurrences of item type 'I' across all 32 subjects, so the frequency rate for that item type was calculated as 107/45 ₃ 2.38, which is 'very high'.

A type/token frequency of 1.32, for example, was rated as ‘very high’ (VH), while a type/token frequency of .03 was rated ‘low’. In addition, this quantitative analysis reflects trends since not all strategies were verbalized or verbalized every time they were used.

d Procedures for qualitative analysis: Once significant relation-ships 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 which 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.

III. RESULTS

While the full set of findings appears in Cohen and Upton (2006), this article will focus on just an illustrative subsample of what the study revealed. Sections 1– 4 will deal with the answer to the first research question concerning the reading and test-taking strategies that respondents reported using for completing the 10 item types. Section 5 will briefly respond to the issue of whether the Reading to Learn and the Inferencing items required and assessed different academic-like approaches to reading than the Basic Comprehension questions, and will also address the issue of item difficulty.

1 Frequency of strategy use across item types

It appeared that the Reading to Learn and the Inferencing items did not require nor assess different academic-like approaches to reading than the Basic Comprehension questions. Table 5 provides the strategy use frequencies across all item types. The accuracy rate (see Table 6) 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. These findings are not consistent with the expectations of the TOEFL committee based on the design principles document. Let us now take a look at the results for representative items (BC-v, I, R2L-ps) from each broad category (Basic Comprehension, Inferencing, and Reading to Learn). The keys in Figure 1 should be referred to when interpreting the different font styles and abbreviations used in the descriptions and examples.

2 Basic comprehension-vocabulary (BC-v)

This item type is intended to ‘measure examinees’ ability to compre-hend the meanings of individual words and phrases as used in the context of the passage’ (ETS, 2003: 4). Examinees need to select the option that can replace the targeted word while preserving the author’s intended meaning in the text context. The accuracy rate for this item type for all 64 attempts was 52/64 ³ 81% (J ³ 81%, C ³ 88%, K ³ 69%, O ³ 88%). Table 7 describes the most frequently used reading and test-taking strategies for this item type.

In reviewing the reading and test-taking strategies for this item type, as given in Table 7, the most notable strategy trends were as follows.

a Strategy Trend One: 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:

1) '**Well, this word** rate. [*Returns to passage*.] **Oh, when they report** positive feelings**, and** rate cartoons, **they become even happier**.' (C5, T1P1Q9)
2) ''' seep seep? **I don't know this word. Let's go to the sentence in the text.**' (K4, T1P3Q7)

Table 5 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
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
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

*Frequency rate = no. of occurrences / no. 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'. Strategies that were used at a low (≤.30) rate across *all* item types were not included in the table.

Table 6 Overall accuracy rate of responses by item type

Item type	Correct attempts	Total attempts	Accuracy rate (%)
BC-v	52	64	81
BC-pr	29	33	88
BC-ss	30	37	81
BC-f	52	64	81
BC-n/e	22	26	85
I	25	45	56
I-rp	23	27	85
I-it	50	55	91
R2L-ps	142	162	88
R2L-st	46	50	92

Figure 1 Keys to font styles and abbreviations used in item type 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 no.	Reading Strategy (e.g., R1 = Reading Strategy 1.)
T no.	Test-Management Strategy (e.g., T1 = Test-Management Strategy 1.)
TW no.	Test-Wiseness Strategy (e.g., TW1 = Test-Wiseness Strategy 1.)
T no. P no. Q no.	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 no.	Specific subject number. J = Japanese subjects, C = Chinese subjects, K = Korean subjects, O = 'Other' subjects (e.g., J1 = Japanese subject no. 1.)
BC	Basic Comprehension item types. There are five different types: BC-v, BC-pr, BC-ss, BC-f, and BC-n/e.
I	Inferencing item types. There are three different types: I, I-rp, and I-it.
R2L	Reading to Learn item types. There are two different types: R2L-ps and R2L-st.

Table 7 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
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

*Strategies are ranked and grouped according to three levels of frequency rate (no. of occurrences/no. of items of that type): very high (≥1.0), high (≥.50), and moderate (≥.30).

b Strategy Trend Two: Two common 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. For example:

3) '**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)

c Strategy Trend Three: *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. Here is an example of strategy T10:

4) '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)

d Strategy Trend Four: 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. The following are examples of strategies T12 and T16:

5) '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)

3 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: 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 this item type for the 45 attempts was 25/45 ³ 56% (J ³ 78%, C ³ 73%, K ³ 31%, O ³ 50%). Table 8 outlines the most frequently used reading and test-taking strategies for this item type.

In reviewing the most common reading and test-taking strategies for this item type, as given in Table 8, the following notable strategy trends emerge:

Table 8 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
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

*Strategies are ranked and grouped according to three levels of frequency rate (no. of occurrences/no. of items of that type): very high (≥1.0), high (≥.50), and moderate (≥.30).

a Strategy Trend One: 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 pas-sage meaning (T28; T22) occurred at very high (2.24) and high (.93) rates, respectively. Examples of strategies T5, T22, and T28:

6) '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)
7) '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)

b Strategy Trend Two: Other strategies reflected the need for participants 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. Here are examples of strategies R6 and R9:

8) [*Rereads paragraph*] 'From the second paragraph, it can be inferred that the higher the mountain is the younger it is.'
(C5, T1P3Q3)
9) [*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)

c Strategy Trend Three: This item type also proved to be among the most difficult of all the item types for subjects to understand. This difficulty is reflected in strategy use. For example, *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. The following are examples of strategies T1, T2, and T3:

10) [*Rereading the question*.] '…<u>inferred</u> 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)
11) 'I guess, my understanding of the question is wrong. None of the options talks about positive things. I'll reread it.**'** (O5, T2P3Q9)

4 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: 15). The R2L-ps items were meant to call on respondents to *read through the entire text* in order to select those three statements which served to describe the text in a summary fashion. The accuracy rate for this item type for the 54 attempts (with three correct choices possible for each item) to the five different items was 142/162 ₃ 88% (J ₃ 87%, C ₃ 88%, K ₃ 82%, O ₃ 93%). Table 9 outlines the most frequently used reading and test-taking strategies for this item type.

In reviewing the reading and test-taking strategies that occurred for this item type, as given in Table 9, the most notable strategy trends included the following.

Table 9 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
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

*Strategies are ranked and grouped according to three levels of frequency rate (no. of occurrences/no. of items of that type): very high (≥1.0), high (≥.50), and moderate (≥.30).

a Strategy Trend One: 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:

12) [*Reads question and introductory sentence. Reads first summary sentence option.*] 'This is correct, so I choose this. [*Reads through Option 2–6, selecting Option 4. After reading Option 6:*] Let's go back to the passage.' (J1, T2P1Q13)
13) [*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)

b Strategy Trend Two: Most of the remaining 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. For example, 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. Here are examples of strategies T22 and T28:

14) [*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 of an individual's emotions which may in turn influence the individual's expressions.' (C8, T1P1Q13)
15) [*Reads Option 4.*] 'No, it doesn't talk about irrigation. It talks about excessive water, so this is not it.' (K2, T2P1Q13)

c Strategy Trend Three: 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, for example: *considering and then postponing consideration of option(s)* (T16) occurred at a very high rate (2.93) and *paraphrasing the meaning of the option* (T14) occurred at a high rate (.57). Examples of strategies T14 and T16 include:

16) [*Reads option 1.*] 'This first sentence means that the two par-ties developed in the process of economic and political com-petition. 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)
17) [*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)

5 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 assess different academic-like approaches to reading than the Basic Comprehension items, since they call for the processing of lengthier texts than in the past, and whether these items were more difficult than the more traditional items.

The study found that the three types of item formats assessed academic reading skills in similar ways, based on the reading and test-taking strategies that the respondents used, and that the new formats were not more

difficult than the more traditional formats, for reasons to be discussed below. A close analysis of strategy use revealed that the R2L and inferencing types of items prompted more noticeable use of those text processing strategies associated with having to cope with longer academic texts – such as greater use of strategies for identifying logical connectors and other markers of cohesion, and determining how sections of passages interrelate, in an effort to establish passage coherence. While the level of focus of the questions (sentence-, paragraph-, or text-level) also proved interesting, we will look here at the strategies deployed and then at the difficulty of the item types.

a The reading and test-taking strategies deployed: While there were some 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.

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-going 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 pas-sage/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, and 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, the expectations of the questions, and the meaning and implications of the different options in light of the text, and then selecting and discarding options based on what they understood about the text.

What set the Reading to Learn items apart 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.

b 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 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 in both content and difficult 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.

In our small-scale study, the R2L item types were not necessarily found to be more difficult for the given respondents, even though the items 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 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.

c 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).

IV. 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*.[6] 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. 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 a 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 participants approached the *new TOEFL* reading section as a test-taking task which required that they perform reading tasks in order to complete them. In other words, the primary goal of

[6] 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 pro-totype 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.

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 abilities. Overall, the findings of the study support 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.

1 The Basic Comprehension item types

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

- 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.
- They challenged the respondents to use powers of inference when they did not recognize vocabulary.
- They called for being mindful of cohesion and coherence issues when attempting to fit the portion of text into the larger discourse.
- 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.
- 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.
- 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.
- The respondents were clearly engaged in problem-solving activities.
- The length of the texts and the nature of the items seemed to pre-clude 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 under-standing 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.

2 The Inferencing item types

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

- 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.

- 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.
- 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.
- 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.
- 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.
- 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.
- 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.

3 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: 6). One of the Reading to Learn formats intends to measure the extent to which L2 readers can complete a 'prose summary' through questions which are referred to as 'multiple-selection multiple-choice responses'. It entails the dragging and dropping of the best descriptive statements about a text.

One can argue whether or not this is truly a summarization task as no writing is called for, and even the set of possible 'main points' is provided for the respondents so that they only need to select those which 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 was found to be a challenging task for Brazilians reading texts and preparing summaries in English as a foreign language, primarily because the respondents had insufficient grasp of the vocabulary to determine what was central and what was more secondary in nature (Cohen, 1994b). 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 under-standing 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 demanding items (e.g., requiring strategies for retaining ideas in working memory, for identifying markers of cohesion, and for perceiving the overall meaning), 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 was somewhat constricted due to repeated 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:

18) [*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 sub-ject 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.

4 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 assure that the four RAs coded the verbal report, as well as their observed and non-verbalized 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 (such as regards the difficulty level of the R2L items) 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 effectiveness of test takers' 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.

V. CONCLUSION

This study set out to determine whether the *new TOEFL* is actually measuring what it purports to measure, as revealed through verbal reports. 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. 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.

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 assessed different academic-like approaches to reading than the Basic Comprehension items. On the basis of findings from this primarily qualitative study, we would contend that the three task formats on the LanguEdge prototypical tests appear to assess similar components of academic reading, as well as test-taking ability. The modifications to the Basic Comprehension item types – such as 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, in fact, made them reflect academic-like tasks which elicit 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 evaluating 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.

Acknowledgements

This study was commissioned and funded by the Educational Testing Service. Special thanks are due to Mary Enright and the ETS TOEFL Committee of Examiners for their assistance in providing access to the resources needed to conduct this study and for their input on the study design, methodology, and final report. A complete description of the study can be found in the ETS Monograph by Cohen and Upton (2006).

VI. REFERENCES

Allan, A. 1992: Development and validation of a scale to measure test-wiseness in EFL/ESL reading test takers. *Language Testing 9(2), pp. 101–22.*

Anderson, N.J. 1991: Individual differences in strategy use in second language reading and testing. *Modern Language Journal 75(4), pp. 460–72.*

Bachman, L.F. and **Palmer, A.S.** 1996: Language testing in practice. Oxford: Oxford University Press.

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

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

Block, E. 1986: The comprehension strategies of second language readers. *TESOL Quarterly 20(3), pp. 463–94*.

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), pp.277–303.*

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

Carrell, P.L. and **Grabe, W.** 2002: Reading. In Schmitt, N., editor, *An introduction to applied linguistics,* London: Arnold, *pp. 233–50*.

Cavalcanti, M.C. 1987: Investigating FL reading performance through pause protocols. In Faerch, C. and Kasper, G., editors, *Introspection in second language research,* Clevedon, UK: Multilingual Matters, *230–50*.

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

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

——1991: Feedback on writing: The use of verbal reports. *Studies in Second Language Acquisition 13(2), pp. 133–59*.

——1994a: *Assessing language ability in the classroom,* second edition. Boston: Newbury House/Heinle and Heinle.

——1994b: English for academic purposes in Brazil: The use of summary tasks. In Hill, C. and Parry, K., editors, *From testing to assessment: English as an international language,* London: Longman, *pp. 174–204*.

——2000: Exploring strategies in test taking: Fine-tuning verbal reports from respondents. In Ekbatani, G. and Pierson, H., editors, *Learner-directed assessment in ESL,* Mahwah, NJ: Erlbaum, *pp. 131–45*.

Cohen, A.D. 2005: *Coming to terms with language learner strategies: What do strategy experts think about the terminology and where would they direct their research?* Working Paper No. 12. Research Paper Series. Auckland, NZ: Centre for Research in International Education, AIS St. Helens. http://www.crie.org.nz/research_paper/Andrew%20Cohen%20WP12.pdf.

Cohen, A.D. and **Aphek, E.** 1979: *Easifying second language learning.* A research report under the auspices of Brandeis University and submitted to the Jacob Hiatt Institute, Jerusalem. Educational Resources Information Center, ERIC ED *pp. 163 753*.

Cohen, A.D. and **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), pp. 63–73*.

——1990: Feedback on compositions: Teacher and student verbal reports. In Kroll, B., editor, *Second language writing: Research insights for the classroom,* Cambridge: Cambridge University Press, *pp. 155–77*.

Cohen, A.D., Glasman, H., Rosenbaum-Cohen, P.-R., Ferrara, J. and **Fine, J.** 1979: Reading English for specialized purposes: Discourse analysis and the use of student informants. *TESOL Quarterly* 13(4), 551–64. Reprinted in Carrell P. L. et al., editors, 1988: *Interactive approaches to second language reading*, Cambridge: Cambridge University Press, *pp. 152–67*.

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

Cohen, A.D. and **Upton, T.A.** (2006): *Strategies in responding to the New TOEFL reading tasks* (TOEFL

Monograph Series Report No. 33). Princeton, NJ: Educational Testing Service. http://www.ets.org/Media/Research/pdf/RR-06-06.pdf
Enright, M.K., Grabe, W., Koda, K., Mosenthal, P., Mulcahy-Ernt, P. and **Schedl, M.** 2000: *TOEFL 2000 reading framework: A working paper* (TOEFL Monograph Series Report No. 17). Princeton, NJ: Educational Testing Service.
Enright, M.K. and **Schedl, M.** April 2000: *Reading for a reason: Using reader purpose to guide test design* (A Draft TOEFL 2000 Report). Princeton, NJ: Educational Testing Service.
Ericsson, K.A. and **Simon, H.A.** 1993: *Protocol analysis: Verbal reports as data*, revised edition. Cambridge, MA: MIT Press.
ETS 2003: *Task specifications for* Next Generation TOEFL *reading test.* Unpublished work. Princeton, NJ: Educational Testing Service.
Fransson, A. 1984: Cramming or understanding? Effects of intrinsic and extrinsic motivation on approach to learning and test performance. In Alderson, J.C. and Urquhart, A.H., editors, *Reading in a foreign language*, London: Longman, *pp. 86–121*.
Grabe, W. 2004: Research on teaching reading. *Annual Review of Applied Linguistics 24, pp. 44–69.*
Green, A.J.F. 1998: *Using verbal protocols in language testing research: A handbook*. Cambridge: Cambridge University Press.
Hosenfeld, C. 1984: Case studies of ninth grade readers. In Alderson, J.C. and Urquhart, A.H., editors, *Reading in a foreign language*, London: Longman, *pp. 231–49*.
Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P., and **Taylor, C.** 1999: *TOEFL 2000 framework: A working paper* (TOEFL Monograph Series Report No. 16). Princeton, NJ: Educational Testing Service.
Jourdenais, R. 2001: Protocol analysis and SLA. In Robinson, P., editor, *Cognition and second language acquisition*, New York: Cambridge University Press, *pp. 354–75*.
Kern, R.G. 1994: The role of mental translation in second language reading. *Studies in Second Language Acquisition 16(4), pp. 441–61*.
Kintsch, W. and **van Dijk, T.A.** 1978: Toward a model of text comprehension and production. *Psychological Review 85(5), pp. 363–94*.
LanguEdge Courseware: handbook for scoring speaking and writing. 2002: Princeton, NJ: Educational Testing Service.
LanguEdge Courseware Score Interpretation Guide. 2002: Princeton, NJ: Educational Testing Service.
Laufer, B. 1991: *Similar lexical forms in interlanguage*. Tübingen, Germany: Gunter Narr.
Leow, R.P. and **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), pp. 35–57.*
Nevo, N. 1989: Test-taking strategies on a multiple-choice test of reading comprehension. *Language Testing 6(2), pp. 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. and **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), pp. 245–50*.
Raimes, A. 1987: Language proficiency, writing ability, and composing strategies: A study of ESL college student writers. *Language Learning 37(3), pp. 439–67*.
Singhal, M. 2001: Reading proficiency, reading strategies, metacognitive awareness and L2 readers. *The Reading Matrix* 1(1) [8 pages].
Skehan, P. 1998: Task-based instruction. *Annual Review of Applied Linguistics 18, pp. 268–86*.
Skibniewski, L. 1990: The writing processes of advanced foreign language learners: Evidence from thinking aloud and behavior protocols. In Fisiak, J., editor, *Papers and studies in contrastive linguistics*, Poznan, Poland: Adam Mikiewicz University, pp. *193–202*.
Stemmer, B. 1991: *What's on a C-test taker's mind: Mental processes in C-test taking.* Bochum: Brockmeyer.

Swales, J. 1981: *Aspects of article introduction*. Birmingham, UK: The University of Aston, Language Studies Unit.
—— 1990: *Genre analysis: English in academic and research settings*. Cambridge: 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 pages].
—— 1998: 'Yuk, the skin of insects!' Tracking sources of errors in second language reading comprehension. *Journal of College Reading and Learning 29(1), pp. 5–20*.
Upton, T.A. and **Lee-Thompson, Li-Chun.** 2001: The role of the first language in second language reading. *Studies in Second Language Acquisition 23(4), pp. 469–95*.
Urquhart, S. and **Weir, C.** 1998: *Reading in a second language: Process, product and practice*. Harlow, Essex: Longman.
Warren, J. 1996: How students pick the right answer: a 'think aloud' study of the French CAT. In Burston, J., Monville-Burston, M. and Warren, J., editors, *Australian Review of Applied Linguistics, Occasional Paper No. 15, pp. 79–94*.
Williams, E. and **Moran, C.** 1989: Reading in a foreign language at intermediate and advanced levels with particular reference to English. *Language Teaching 22(4), pp. 217–28*.
Zamel, V. 1983: The composing processes of advanced ESL students: Six case studies. *TESOL Quarterly* 17(2), 165–87.

Appendix A: Item type descriptions

Descriptions are all taken from *ETS*, 2003, unless otherwise noted.

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).

Vocabulary Items (BC-v)
'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).

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

Sentence Simplification Items (BC-ss)
'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).

Factual Information Items (BC-f)
'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).

Negative Fact Items (also called Not/Except Items) (BC-n/e)
'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).

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).

Inferencing Items (I)
'These items measure examinees' ability to comprehend an argument or an idea that is strongly implied but not explicitly stated in the text' (p. 25).

Rhetorical Purpose Item (I-rp)
'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).

Insert Text Items (I-it)
'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).

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 relation-ships, compare-contrast relationships, arguments, etc.

Prose Summary Items (R2L-ps)

'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.***

Answer choices

- The fine arts are only affected by the laws of physics because of the limitations of the materials that are used.
- Applied-art objects are bound by the laws of physics in two ways: by the materials used to make them, and the function they are to serve.
- Crafts are known as 'applied arts' because it used to be common to think of them in terms of their function.
- In the fine arts, artists must work to overcome the limitations of their materials, but in the applied arts, artists work in concert with their materials.
- Making fine-art objects stable requires an understanding of the properties of mass, weight, distribution, and stress.
- In the twentieth century, artists working in the fine arts often treat materials in new ways whereas applied arts specialists continue to think of crafts in terms of function.

Schematic Table Items (R2L-st)

'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' (*LanguEdge Courseware*, 2002: 48).