Language Learning 56:4, December 2006, pp. 633–670

ESL Reading Strategies: Differences in Arabic and Mandarin Speaker Test Performance

Marilyn L. Abbott
Alberta Education

This study was undertaken to test the hypothesis that reading comprehension items, which elicit specific bottom-up and top-down strategies, favor certain linguistic/cultural groups. Verbal report data were collected from Arabic- and Mandarin-speaking English as a second language (ESL) learners to identify the reading strategies involved in carrying out 32 reading questions. Then a confirmatory approach to differential item functioning was used to determine whether bottom-up and top-down items functioned differentially for equal-ability Arabic and Mandarin ESL learners. Results revealed systematic group performance differences in four bottom-up and three top-down strategy categories. Items involving breaking a word into smaller parts, scanning, paraphrasing, and matching were found to favor Mandarin speakers, whereas items

I wish to thank Tracey Derwing and the Canadian Centre for Language Training and Assessment for their support of this project. I would also like to thank the three anonymous reviewers for their helpful comments.

Funding for this research has been provided by an Isaac Walton Killam Memorial Scholarship, a Social Sciences and Humanities Research Council of Canada Doctoral Fellowship, the Prairie Centre of Excellence for Research on Immigration and Integration, and a TOEFL Small Grant for Doctoral Research in Second Language Assessment.

Correspondence concerning this article should be addressed to Marilyn Abbott, Alberta Education, 44 Capital Boulevard, 10044-108 Street, Edmonton, Alberta, Canada, T5J 5E4. Internet: Marilyn.Abbott@gov.ab.ca

involving skimming, connecting, and inferring were found to favor Arabic speakers.

The Canadian federal government provides language training to immigrants who have limited or no proficiency in an official language on arrival. Many immigrants, however, are unable to access more than the federally funded maximum of 1,500 hr of instructional support. Consequently, it is necessary that immigrants' English language levels be accurately assessed so that they can be placed in the most appropriate levels of instruction.

It is also crucial to ensure that placement and proficiency tests provide equal opportunities for all immigrants to demonstrate their abilities relevant to "the construct(s) the test is intended to measure" (*Standards for Educational and Psychological Testing*, 1999, p. 74). For example, if a reading comprehension test is made up of question types that elicit strategies that are well developed in one specific linguistic/cultural group but not in another, then the assessment might unfairly favor the first group over the second. In other words, if the questions involve reading strategies that are more familiar to members of one language or cultural group, then the assessment might be easier for individuals of that group.

Since its inception in 1996, the Canadian Language Benchmarks Assessment (CLBA) has predominantly been used to assess the English language skills of newcomers to Canada. The CLBA is promoted as a task-based tool (i.e., it includes a range of tasks of different types) designed to assess language proficiency in the areas of listening, speaking, reading, and writing. Initially, the main purpose of the assessment was to determine newcomers' entry points in English as a second language (ESL) programs. Currently, the CLBA is also being used as a means of establishing admissible levels of English language proficiency in some postsecondary institutions. This usage has moved the CLBA into the realm of high-stakes testing.

To date, the extent to which the CLBA reading items might favor examinees from particular language or cultural groups has

not been the focus of any empirical research. In an attempt to fill this void and to extend our understanding of cross-cultural reading strategy use, the purpose of this study was to (a) develop a bottom-up, top-down reading strategy framework designed to evaluate whether the CLBA Reading Assessment produces comparable results for equal-ability Arabic and Mandarin first language speakers (L1) and (b) to identify groups of items that elicit systematic performance differences for Arabic- and Mandarin-speaking examinees on the CLBA Reading Assessment. Arabic and Mandarin ESL learners were selected for three main reasons: first, they are currently two of the largest recent immigrant groups in Canada; second, both languages are radically different from English and from each other in terms of orthographic script; and third, the two groups are culturally distinct.

Although a plethora of questionnaire research results indicate that cultural background affects second language (L2) learning strategy selection and use (e.g., Bedell & Oxford, 1996; Harshbarger, Ross, Tafoya, & Via, 1986; Levine, Reves, & Leaver, 1996; Reid, 1995; Willing, 1988), few studies have specifically focused on how L2 reading strategies interact with L1 and cultural background to affect test performance. It has been determined, however, that ESL reading comprehension tests often focus on low-level linguistic cues, which tend to reward bottom-up as opposed to top-down reading strategies (Hill & Parry, 1989, 1992; Purpura, 1997). Bottom-up reading comprehension strategies are data-driven, detail-oriented strategies, whereas top-down strategies are conceptually driven, big-picture oriented strategies (Carrell, 1983). Parry (1996) found that when attempting English academic reading tasks, different cultural groups use strikingly different reading strategies that she argues are related to their different language backgrounds and different experiences of literacy. For example, whereas Chinese students showed a definite preference for bottom-up methods, Nigerian students reported a strong tendency to use top-down strategies. In another crosslinguistic study of ESL reading, Fender (2003) discovered that native Arabic ESL learners were more accurate in comprehending

and integrating words into larger phrase and clause units than Japanese ESL learners. This suggests that Arabic ESL learners might have a proclivity for using top-down reading strategies.

Although Durgunoglu and Hancin (1992) view the top-down, bottom-up models of reading as being outdated in the realm of L1 research (which currently emphasizes the importance of visual processing), some L2 reading researchers (e.g., Brantmeier, 2000, 2003a, 2003b; Coady, 1997; Eskey, 1997; Liontas, 1999; Parry, 1996; Schueller, 2004; Storey, 1997; Young & Oxford, 1997) currently rely largely on bottom-up, top-down models of reading comprehension. Therefore, the substantive and statistical analyses conducted in this study were also based on a bottom-up, top-down approach to reading. This allowed for an evaluation of the appropriateness of using a bottom-up, top-down reading strategy framework for modeling ESL reading comprehension.

In the following literature review, I examine the ESL reading strategies literature and consider the effects of culture, education, and language on the development and use of reading strategies. Then I present some of the issues in assessing reading and discuss the differential item functioning (DIF) studies of multiple-choice ESL language proficiency and placement tests. Finally, I apply a DIF analysis framework to the study of Arabic and Mandarin speaker performance differences in ESL reading strategy use on the CLBA Reading Assessment.

Literature Review

Reading Comprehension Strategies

Researching L2 reading comprehension strategies has proved to be a complex endeavor, as the concept of strategy is difficult to define, observe, measure, describe, and classify. Despite the lack of consensus regarding what constitutes a strategy, numerous researchers use the term *strategies* to refer to the mental

processes or behaviors that language learners employ in L2 acquisition, L2 use, or L2 testing situations (Alderson, 1984; Cohen, 1998; Hosenfeld, 1977; O'Malley & Chamot, 1990; Oxford, 1990; Purpura, 1997). According to Cohen, language use and test-taking strategies are the "mental operations or processes that learners consciously select when accomplishing language tasks" (p. 92). By adapting this definition to the context of reading, reading comprehension strategies might be defined as the mental operations or comprehension processes that readers select and apply in order to make sense of what they read. Because strategies are generally considered to be conscious or at least potentially conscious, they are open to inspection (Weinstein & Mayer, 1986).

Examples of some commonly identified reading strategies are skimming for gist, scanning for details, guessing, recognizing cognates and word families, predicting, activating general knowledge, making inferences, following references, and separating main ideas from supporting ideas (Barnett, 1988). Although some reading experts (Davis, 1968; Drum, Calfee, & Cook, 1981; Munby, 1978) classify these strategies as reading skills, microskills, or subskills, others (Alexander & Jetton, 2000; Carrell, 1989; Duffy et al., 1987; Robb, 1996) refer to these behaviors as strategies, as they assume that a reading skill becomes a strategy when the reader can use it independently, reflect on it, and understand what it is, how it works, and when to apply it to new texts. This assumption will be adopted in the current study.

Examples of Bottom-up, Local and Top-down, Global Reading Strategies

Examples of bottom-up, local, language-based reading strategies that focus primarily on word meaning, sentence syntax, or text details and are associated with attending to lower level cues are as follows:

1. breaking words into smaller parts
2. using knowledge of syntactic structures or punctuation

3. scanning for specific details
4. paraphrasing or rewording the original text
5. looking for key vocabulary or phrases

Some top-down, global, knowledge-based reading strategies that focus primarily on text gist, background knowledge, or discourse organization and are associated with attending to higher level cues include the following:

1. recognizing the main idea
2. integrating scattered information
3. drawing an inference
4. predicting what might happen in a related scenario
5. recognizing text structure

These strategies appear in standard classifications employed in one or more of the following studies: Anderson, 1991; Block, 1986; Carrell, 1989; Phakiti, 2003; Pritchard, 1990; Purpura, 1997; Schueller, 2004; and Young and Oxford, 1997.

Influence of Culture, Education, and Language on the Acquisition of EFL/ESL

Although some cultural and educational factors have been shown to influence strategy preferences (e.g., Bedell & Oxford, 1996; Harshbarger et al., 1986; Levine et al., 1996; Pritchard, 1990; Reid, 1995; Willing, 1988), little explanation has been provided as to why this occurs. Thus, in this subsection, an attempt will be made to explain why intermediate Arabic and Chinese ESL learners tend to use different reading strategies.

Instructors of reading in English influence the way their students approach text by teaching them to read in particular ways. For example, it is often cited that Chinese teachers tend to use traditional teacher-centered approaches to teaching English as a foreign language (EFL) (Burnaby & Sun, 1989; Parry, 1996;

Penner, 1995). As a result, Chinese EFL learners are taught to pay close attention to word-level cues (i.e., morphology and syntax). According to Fischer-Kohn (1986, cited in Kohn, 1992), Chinese teachers of reading in English encourage their students to do the following:

1. read slowly and take care that they know each word as they go
2. vocalize or voice the material, either loudly or silently
3. reread difficult sentences until they are understood
4. look up definitions of all unknown words in a dictionary
5. analyze complex structures carefully (p. 121)

Thus, it appears that Chinese EFL learners are taught to use bottom-up strategies, as they are expected to carefully scrutinize each word in the text and memorize grammar rules and exceptions (Kohn).

In contrast, the general trend in Arab nations is to place more emphasis on student-centered EFL activities that encourage linguistic interaction through the use of authentic, real-life tasks (Kharma, 1998). These types of communicative activity focus on developing functional language skills in a learning environment that stresses meaning over form. As Parry (1996) suggested, authentic reading activities that emphasize reading for meaning tend to encourage a global, top-down approach to text. Therefore, it is likely that the exposure that Arab EFL students receive to communicative activities promotes the development of top-down reading strategies.

Effects of Linguistic Differences on the Acquisition of EFL/ESL

Research suggests that language-specific differences are related to differences in processing skills and strategies in reading (Akamatsu, 2003; Chen, 1992; Fender, 2003; Koda, 1988). For example, as mentioned in the introduction, in a crosslinguistic

study of ESL reading skills, Fender found that Arabic ESL learners were more accurate in comprehending and integrating words into larger phrase and clause units than the Japanese ESL learners in the study. Japanese (kanji), like Chinese, uses an orthography that encodes language at the level of morphemes, which, in general, correspond to words and affixes (Chen). Therefore, one might hypothesize that Chinese ESL learners would also have difficulty with word integration when reading in English.

According to Abu-Rabia (1997), "Arabic is perhaps the only language in the world in which readers must first understand the sentence in order to recognize the word" (p. 76). Because the vowels are not represented in Arabic orthography, Arabs might be less dependent on local cues in the printed word when reading. If reading in Arabic encourages a reliance on higher level contextual cues and strategies, it is possible that the Arabic ESL learners in Fender's (2003) study were more successful integrators than the Japanese ESL learners because they effectively transferred their well-developed L1 reading strategies to the L2 reading task. It is likely that the reduction of the extent of Arab readers' dependence on the visual stimulus causes them to develop more effective top-down reading comprehension processes. As a result, when processing printed material in English, Arabic ESL learners might rely more on their background and contextual knowledge than on their linguistic knowledge and, consequently, have a proclivity for using top-down reading strategies over bottom-up ones. On the contrary, it is possible that the careful approach that Chinese ESL learners take might cause them to be distracted by less relevant textual information and, as a result, they might not be as skilled at integrating words into larger units.

Native speakers of Chinese, however, develop a sophisticated set of orthographic processing skills through their literacy experiences. When compared with printed words in alphabetic (e.g., English) or consonantal (e.g., Arabic) orthographies, Chinese encodes morphemes with much less phonology (Akamatsu,

1999). Consequently, whereas Chinese word recognition requires extensive orthographic processing skills, alphabetic or consonantal orthographies require a greater connection to phonemes and phonology. Therefore, Chinese ESL learners might be able to utilize their L1-based processing skills to develop a set of graphic ESL word representations that facilitate English word processing (Akamatsu, 2003).

Although L1 Arabic literacy skills are developed through reliance on phonological processing skills, as Arabic orthography has a highly consistent set of grapheme-phoneme correspondences, more mature readers must learn to use an orthography that does not include diacritic marks that signal the vowels (Abu-Rabia, 1999). In comparison, reading in English encourages greater reliance on (a) phonological skills for decoding words with regular grapheme-phoneme correspondences and (b) orthographic processing skills for decoding words with grapheme-phoneme irregularities (e.g., business, cough, iron; Katz & Frost, 1992). Therefore, it is likely that the Arabic and Chinese ESL learners' primary L1 processing skills and strategies that have been developed through exposure to distinct languages and literacy practices will differentially influence the development of their ESL processing skills and strategies.

Reading Assessment

In order to make inferences about examinee reading competency from test performance, the reading skills and strategies assessed in the items must adequately represent and be relevant to the construct of reading comprehension.

> Invalidly low scores should not occur because the assessment is missing something relevant to the focal construct that, if present, would have permitted the affected persons to display their competence. Moreover, invalidly low scores should not occur because the measurement contains something irrelevant that interferes with the affected person's demonstration of competence. (Messick, 1996, p. 252)

If a test predominantly measures low-level linguistic skills and strategies, invalidly high scores might be attained by students well prepared on the represented skills or processes that are measured by the reading items but ill-prepared on the underrepresented ones. Therefore, because any test is merely a sample of the underlying construct, it is important to identify the relevant skills and strategies that the items are assumed to assess. If the construct is adequately represented and the tasks are relevant to the target language use domains, then the test can be used to predict real-life reading ability and identify readers' strengths and weaknesses.

Reading assessment researchers are not only concerned with how well reading performance on a test predicts how examinees will read in other real-world settings; they are also interested in discovering which cognitive characteristics and background variables influence test performance (Alderson, 2000). Thus, reading comprehension assessments that seek to support instructional decision-making for language learners from different cultural/linguistic backgrounds must take into consideration how cultural/linguistic differences affect test performance. Comparing examinees' scores from different cultural/linguistic groups on reading items and identifying certain patterns of correct and incorrect responses is a first step in determining the cognitive factors and/or strategies that might affect reader comprehension.

Differential Item Functioning

Differential item functioning (DIF) is present when examinees from distinct groups have different probabilities of answering an item correctly after controlling for overall test performance (Shepard, Camilli, & Averill, 1981). DIF methods match examinees on ability (usually total test score) to determine whether comparable examinees from different populations perform the same on individual items or groups of items. For example, one would expect Arabic- and Mandarin-speaking examinees who have the same total test score to perform in an equivalent manner

on each reading comprehension item. When comparable examinees do not perform the same on specific test items, the items are said to display DIF. Large DIF indexes signify that the items are measuring secondary dimensions that might either be relevant or irrelevant to the construct measured by the test (Shealy & Stout, 1993). If an item is measuring a secondary dimension that is an appropriate part of the intended construct, the secondary dimension is considered auxiliary. Thus, the DIF between groups reflects a true difference in the construct and is considered benign. Alternatively, if an item is measuring an unintended secondary dimension, the secondary dimension is considered nuisance. DIF caused by nuisance dimensions reflects bias that might be thought of as systematic error that distorts the meaning of test inferences for members of a specific group and, therefore, poses a considerable threat to validity (Camilli & Shepard, 1994).

Much of the research regarding the effects of language background on L2 test performance has been concerned with whether EFL/ESL language proficiency and placement tests measure the same constructs for different language groups (e.g., Ackerman, Simpson, & de la Torre, 2000; Brown, 1999; Ginther & Stevens, 1998; Kunnan, 1994). Only a few studies have examined how different language groups perform on multiple-choice proficiency and placement tests at the item level (see Chen & Henning, 1985; Ryan & Bachman, 1992; Sasaki, 1991).

Although the statistical methods utilized in these three DIF studies were relatively useful for flagging DIF items, to understand the nature of DIF, content analyses are also required to determine why the items functioned differentially between the groups. However, the researchers either provided no explanations for the causes of DIF or their attempts at identifying the causes of DIF in many of the items using content analyses were not very successful. For example, of the 22 DIF items identified by Scheuneman's (1979) chi-square method in Sasaki's (1991) study, only 4 of the items had interpretable sources of DIF. Because attempts to understand the "underlying causes of DIF using substantive analyses of statistically identified items have, with few

exceptions, met with overwhelming failure" (Roussos & Stout, 1996, p. 360), Roussos and Stout (1996) proposed a confirmatory approach to DIF. This approach, which is based on the Shealy-Stout multidimensional model for DIF (Shealy & Stout, 1993), was used in the current study.

A Confirmatory Approach to DIF

The Roussos-Stout (1996) approach to DIF is a two-stage approach designed to link substantive and statistical methods in a DIF analysis framework. In the first stage of this framework, DIF hypotheses are generated from theory and substantive item analyses are conducted to classify the items according to an *organizing principal* or theoretical framework. A DIF hypothesis specifies whether an item or group/bundle of items designed to measure the primary or intended dimension also measures a secondary dimension or unexpected dimension that is suspected of producing DIF. The second stage in the DIF analysis framework involves statistically testing the hypotheses generated in stage one. The statistical procedure selected for testing the hypotheses in this study was Simultaneous Item Bias Test (SIBTEST) (Stout & Roussos, 1999).

The SIBTEST quantifies the size of DIF by estimating a measure of the effect size ($\hat{\beta}_{UNI}$) for each item and/or bundle (Stout & Roussos, 1995). $\hat{\beta}_{UNI}$ is interpreted as the expected advantage in proportion or number correct score that one group of examinees has over the other group of examinees on an item or a specific bundle of items. The advantaged group is referred to as the reference group and the disadvantaged group is referred to as the focal group. $\hat{\beta}_{UNI}$ is calculated as the weighted sum of the differences between the proportion-correct or number-correct true scores on the studied item or bundle for examinees in the two groups across all score levels. The true scores are estimated using a regression correction described in Shealy and Stout (1993). The weighted mean difference between the reference and focal groups

on the studied subtest item or bundle across the k subgroups is given by

$$\hat{\beta}_{\text{UNI}} = \sum_{k=0}^{k} p_k d_k$$

where p_k is the proportion of focal group examinees in subgroup k and d_k is the difference in the adjusted means on the studied subtest item or bundle of items for the reference and focal groups, respectively, in each subgroup k. For large samples, $\hat{\beta}_{\text{UNI}}$ has a standard normal distribution with a mean of zero and standard deviation of 1 under the null hypothesis of no DIF. The statistical hypothesis tested by SIBTEST is

$$H_0\colon \beta_{\text{UNI}} = 0$$

versus

$$H_1\colon \beta_{\text{UNI}} \neq 0.$$

The SIBTEST yields the following test statistic for evaluating the $\hat{\beta}_{\text{UNI}}$ null hypothesis of no DIF:

$$\text{SIB} = \frac{\hat{\beta}_{\text{UNI}}}{\hat{\sigma}(\hat{\beta}_{\text{UNI}})}$$

where $\hat{\sigma}(\hat{\beta}_{\text{UNI}})$ is the estimated standard error of $\hat{\beta}_{\text{UNI}}$. SIB is evaluated against the standard normal distribution. A null hypothesis of no DIF is rejected whenever $|\text{SIB}| > z_{1 - {}^1/_2\alpha}$. A statistically significant value of $\hat{\beta}_{\text{UNI}}$ that is positive indicates DIF against the focal group and a negative value indicates DIF against the reference group. See Shealy and Stout (1993) for a comprehensive and technical discussion of the SIBTEST procedure.

In the current study, the following DIF bundle hypotheses were tested using item-level reading response data collected from

250 Mandarin speakers' and 250 Arabic speakers' CLBA Reading Assessments:

1. Arabic-speaking ESL learners will outperform the Mandarin-speaking ESL learners on the bundles of items that rely on top-down processing strategies.

2. Mandarin-speaking ESL learners will outperform the Arabic-speaking ESL learners on the bundles of items that rely on bottom-up processing strategies.

Method

Overview

Verbal report data were collected from seven Arabic- and eight Mandarin-speaking intermediate ESL learners as they completed Form 1, Stage II of the CLBA Reading Assessment. The reading strategies identified in the verbal report data were used to develop the bottom-up, top-down reading strategies framework that was then used in the DIF analyses to evaluate whether the CLBA Reading Assessment produces comparable results for Arabic and Mandarin L1 speakers.

Arabic- and Mandarin-speaking immigrants were recruited from intermediate ESL college classes. Only those students who (a) were literate in their L1 (i.e., they had at least 11 years of basic education in their country of origin), (b) had reached a language threshold in English that would allow them to complete Stage II of the CLBA Reading assessment (i.e., they had mastered the basic vocabulary and decoding skills required for placement in an intermediate ESL class), and (c) had not resided in Canada for more than 2 years were selected.

Materials

The CLBA Reading Assessment is divided into two stages and there are four parallel forms for each stage. Only Form 1,

Stage II was analyzed in this study for two main reasons. First, I conducted a preliminary content analysis that revealed that Stage I of the CLBA Reading Assessment is primarily comprised of bottom-up questions that mainly test vocabulary knowledge and transcoding rather than reading comprehension, whereas Stage II elicits a wider variety of bottom-up and top-down strategies. Previously described inventories of reading processing strategies (see Anderson, 1991; Block, 1986; Carrell, 1989; Phakiti, 2003; Pritchard, 1990; Purpura, 1997; Schueller, 2004; Young & Oxford, 1997) were used as the starting point in this preliminary classification of the items. The second reason that I selected Form 1, Stage II was that the minimum sample size requirement ($n = 250$ in each group) for the DIF analyses was only satisfied by the Form 1 data.

Form 1, Stage II of the CLBA Reading Assessment consists of 8 dichotomously scored constructed-response items and 24 multiple-choice, four-option items. The items follow four passages (Tasks A–D) that represent four different genres and range in length from 251 to 547 words. Because the CLBA is a secured test with copyright limitations, the reading passages and items cannot be released. However, to provide examples for the reader, a reading passage of similar difficulty with corresponding multiple-choice and constructed-response items is included in the Appendix.

Procedures

Verbal report procedures. The verbal report procedures applied in this study follow the initial model suggested by Ericsson and Simon (1993), which was further refined by Pressley and Afflerbach (1995) and then again by Green (1998). Initially, each participant was introduced to the verbal report procedures and provided with a chance to practice his/her verbal reporting skills with four or more reading comprehension questions. The verbal reports were conducted individually in the student's language of choice (i.e., his/her L1, English,

or both languages). Each participant reported in detail what he/she was thinking and what information he/she was attending to when answering each question. If the participants remained silent for more than 5 s, they were reminded to keep talking.

After each verbal protocol was collected, translated into English where necessary, and transcribed, it was segmented and coded by the researcher for types of bottom-up and top-down reading strategy. Each segment of the protocol corresponded to a statement or phrase associated with each strategy that the reader employed. Strategies were defined as each separate action the reader took to process the reading comprehension question and to formulate an answer. The strategy segments comprised the units for analysis. Each segment was assigned one code. Those segments that could not be unambiguously coded were assigned a miscellaneous code.

The sample size for the verbal protocols was determined by data saturation or redundancy. Saturation occurs when no new or useful information about the categories can be obtained (Glaser, 1992; Glaser & Strauss, 1967). Researchers suggested that data saturation is typically reached after the analysis of 5–10 protocols (Glaser & Strauss; Rennie, 1984). In qualitative research, there are no published guidelines or tests of adequacy for estimating the sample size required to reach saturation equivalent to those formulas used in quantitative research. Rather, in qualitative research, the signals of saturation are determined by investigator proclamation and by evaluating the adequacy and the comprehensiveness of the results (Gubrium, 1995). To clarify and elaborate on the reading strategies framework, sampling continued until all properties and dimensions of the categories were identified and no new or relevant data emerged from the participants' verbal reports in either of the language groups. Thus, data collection and analysis occurred concurrently.

After each new participant's data were collected, translated into English where necessary, transcribed, segmented, and

coded, the reading strategy coding schema was revised and the previously coded protocols were recoded using the modified strategy classification schema. Although it appeared that saturation of the reading strategy categories was complete after five Mandarin participants' and four Arabic participants' verbal reports had been coded, to ensure that data saturation had been achieved, verbal reports were collected from three additional Arabic participants and three additional Mandarin participants. Although these last reports did not provide any new reading strategy categories, in some instances they provided clearer examples of the already identified bottom-up and top-down reading strategies. However, these reports did not result in any changes to the bottom-up, top-down reading strategy classification schema.

Consistency of the coding was investigated by having an independent rater code 11 of the 32 CLBA reading items from each of the protocols (34.4% of the total sample). Eleven questions were selected, as they were believed to elicit the full range of bottom-up and top-down strategies identified in the verbal reports. One rater who had no experience with the study was trained to use the coding schema to classify the strategies in the verbal report data. First, the coding schema was discussed with the rater. Next, verbal reports from three randomly selected participants were coded for practice (with the exception of the 11 items used to calculate interrater agreement). Finally, interrater agreement was calculated by assessing the extent to which the researcher and the rater agreed on the codes assigned to each segment within the protocols. This was reported as the percentage of instances where agreement was reached. Of the 456 segments coded, 413 agreements occurred. Therefore, the percentage of total agreement was 90.6%, indicating that the reading strategy segments were consistently coded.

DIF procedures. Comprehensive analyses of the Form 1, Stage II CLBA reading items were conducted using both substantive and statistical methods. Three ESL experts conducted

the substantive analyses of the reading items. The expert judges were asked to classify each of the 32 items into one of the strategy categories that emerged from the verbal report data. After participating in a training session that introduced the judges to the coding schema, they were asked to independently classify the questions according to the reading strategy that they believed was most instrumental in arriving at the answer for each of the CLBA reading items (i.e., each item was classified according to the "most salient" strategy believed to be elicited by the item). Once the judges had finished coding all the items, a meeting was held so that they could reach 100% consensus on the item codings on which they disagreed.

For the statistical analyses, item-level Form 1, Stage II CLBA Reading Assessment response data were collected from 250 previously administered examinee test forms in each of the first language groups (i.e., from 250 Mandarin speakers' and 250 Arabic speakers' assessments). To ensure that the item response data were collected from examinees who had similar levels of L1 literacy and English proficiency as the participants in the verbal report sample, the same selection criteria were used: Data were only collected from examinees who (a) had at least 11 years of basic education in their country of origin, (b) had reached a language threshold in English (i.e., had mastered the basic vocabulary and decoding skills required for moving on to Stage II of the CLBA), and (c) had not resided in Canada for more than 2 years at the time of testing.

Analysis

A bundle DIF hypothesis for each of the reading strategy categories was tested at an alpha level of 0.05 using the computer program SIBTEST (Stout & Roussos, 1999). The alpha value was not adjusted using the Bonferroni correction because the sample size ($n = 250$ per group) was the smallest recommended sample size for conducting reasonably reliable DIF analyses; the smaller the sample size, the larger the standard error of $\hat{\beta}_{\mathrm{UNI}}$, which is

the denominator of the test statistic SIB, and the less likelihood there is of reaching statistical significance.

SIBTEST matches examinees on ability to determine whether comparable examinees from different populations perform the same on individual items or groups of items. The 32 CLBA reading items were grouped into bundles of items based on the judges' consensus codes, and DIF analyses were conducted using the item-level response data to determine whether groups of CLBA items (classified according to the bottom-up, top-down reading strategy framework) functioned differentially for equal-ability Arabic- and Mandarin-speaking ESL learners. These analyses were conducted to show whether there were systematic ways in which the two language/cultural groups responded to groups of test items that were presumed to measure the secondary dimensions of bottom-up and top-down reading strategies.

Results and Discussion

Verbal Report

The verbal report data proved extremely valuable in revealing the reading strategies elicited by the CLBA reading items. The bottom-up, top-down reading strategy classification schema that emerged from the verbal report data is presented in Table 1. Analyses of the verbal protocols identified 12 main reading strategy categories consistent with those identified in the literature (see Anderson, 1991; Block, 1986; Carrell, 1989; Phakiti, 2003; Pritchard, 1990; Purpura, 1997; Schueller, 2004; Young & Oxford, 1997). The strategies include the following: B1, breaking lexical items into smaller parts; B2, scanning for details; B3, identifying synonyms or paraphrases; B4, matching key words to key visuals; B5, matching key vocabulary in the text to key vocabulary in the item; B6, using knowledge of grammar or punctuation; B7, using local context cues to interpret a word or phrase; T1, skimming for gist; T2, connecting or relating information presented

Table 1

Reading strategies used when answering the CLBA reading items

Strategy	Definition
Bottom-up, local strategies	*The reader:*
B1. Breaks lexical items into parts	Breaks words into smaller units to promote comprehension
B2. Scans for explicit information requested in the item	Scans the text for specific details or explicitly stated information requested in the item
B3. Identifies a synonym or a paraphrase of the literal meaning of a word, phrase, or sentence	Identifies or formulates a synonym or a paraphrase of the literal meaning of a word, phrase, or sentence in the text to help answer the question
B4. Relates verbal information to accompanying visuals	Matches verbal information in the text to visual information in the item to answer the question
B5. Matches key vocabulary in the item to key vocabulary in the text	Matches key vocabulary or phrases in the item or options to key vocabulary or phrases in the text
B6. Uses knowledge of grammar or punctuation	Uses awareness of grammar, syntax, parts of speech, or punctuation to help answer the question
B7. Uses local context cues to interpret a word or phrase	Uses the words in a sentence that precede or follow a specific word or phrase to understand a particular word or phrase
Top-down, global strategies	*The reader:*
T1. Skims for gist/identifies the main idea, theme, or concept	Draws on the major points of the passage to answer the question; summarizes main concept
T2. Connects or relates information presented in different sentences or parts of the text	Relates new information to previously stated information to help answer the question; synthesizes scattered information
T3. Draws an inference based on information presented in the text	Makes an inference, draws a conclusion, or forms a hypothesis based on information not explicitly stated in the text to answer the question
T4. Speculates beyond the text	Uses background knowledge to speculate beyond the text
T5. Recognizes discourse format	Uses discourse format or text organization to answer the question (e.g., discriminates between fact and opinion or cause and effect; or notes how the information is presented)

in different parts of the text; T3, drawing an inference based on information presented in the text; T4, using background knowledge to speculate beyond the text; and T5, recognizing discourse format. Extended definitions of each of these reading strategies are presented in Table 1.

Psychometric Characteristics of the CLBA Reading Assessment

Table 2 contains the psychometric characteristics for the CLBA Reading Assessment. The mean total test scores demonstrate that the Mandarin-speaking examinees outperformed the Arabic examinees on the CLBA Reading Assessment by approximately 3.0% on average. A test of the difference in means was significant, $t(498) = -2.031, p < .05$. However, the effect size was small, $d = 0.13$. The results indicate that overall item difficulty and discrimination were comparable across groups.

Differential Item Functioning

The results for the DIF hypothesis tests are reported in Table 3. A bundle $\hat{\beta}_{UNI}$ is interpreted as the expected advantage in the number-correct score that one group has over the other on

Table 2

Descriptive statistics for Form 1 Stage II of the CLBA Reading Assessment

Characteristic	Arabic	Mandarin
No. of Examinees	250	250
No. of Items	32	32
Mean	17.20	18.29
Standard deviation	5.87	6.20
Mean item difficulty	0.54	0.57
Mean item discrimination[a]	0.40	0.42

[a]Point biserial correlation

Table 3

Differential bundle functioning results

Bundle	No. of CLBA reading items	$\hat{\beta}_{UNI}$	Favors
Bottom-up			
B1: Lexical	1	0.176*	Mandarin
B2: Scanning	5	0.233*	Mandarin
B3: Synonym/paraphrase	7	0.915*	Mandarin
B4: Visuals	1	−0.048	ns[a]
B5: Matching key words	4	0.578*	Mandarin
Top-down			
T1: Skimming	1	0.137*	Arabic
T2: Connecting	3	0.463*	Arabic
T3: Inferencing	8	0.445*	Arabic
T4: Speculating	1	0.044	ns
T5: Discourse format	1	−0.033	ns

[a]Not significant.
*$p < .05$.

a specific bundle of items. Therefore, in Table 3, the $\hat{\beta}_{UNI}$ estimate of 0.915 for bundle B3—identifying synonyms or paraphrases—means that a randomly chosen Mandarin-speaking examinee can be expected to obtain a number-correct score of approximately one score point more on the seven B3 items than an Arabic-speaking examinee of equal reading ability. In the case that a bundle only has one item, $\hat{\beta}_{UNI}$ is interpreted as the expected advantage in proportion-correct that one group has over the other on the studied item. As a result, Arabic-speaking examinees can be expected to obtain approximately one tenth of a score point more than equal-ability Mandarin-speaking examinees on the item in bundle T1—skimming for gist. Although this difference is not large, if other similar items were added to the test, this would create a further disadvantage for the Mandarin speakers.

Because the expert judges did not use two of the codes (B6 and B7) as the "most salient" code for any of the items, these two

strategy categories were deleted from the DIF analyses. Consistent with expectations based on the reading strategy framework, 7 of the 10 bundle DIF hypotheses that were tested were supported by the SIBTEST analyses. Systematic group differences were found in four of the bottom-up strategy categories and were found to favor the Mandarin speakers: B1, breaking lexical items into smaller parts; B2, scanning for details; B3, identifying synonyms or paraphrases; and B5, matching key vocabulary in the text to key vocabulary in the item. Therefore, on the CLBA Reading Assessment, Arabic-speaking examinees were found to differ systematically from Mandarin-speaking examinees with comparable top-down reading scores on their skill in answering items presumed to elicit these four strategies (i.e., B1, B2, B3, and B5). Because the items in these bundles have a strong focus on word-level strategies, which utilize knowledge of linguistic parts and forms to interpret text on a word-by-word basis, this factor might have contributed to DIF favoring the Mandarin speakers.

Systematic group differences were also found for three of the five top-down strategy categories and were found to favor the Arabic speakers: T1, skimming for gist; T2, connecting or relating information presented in different parts of the text; and T3, drawing an inference based on information presented in the text. This means that on the CLBA Reading Assessment, Mandarin-speaking examinees were found to differ systematically from Arabic-speaking examinees with comparable bottom-up reading scores on their skill in answering items presumed to elicit these three top-down strategies (i.e., T1, T2, and T3). Because the items in these bundles have a strong focus on semantic strategies that utilize global contextual information and existing knowledge of real-life situations and discourse knowledge, this factor might have contributed to DIF favoring the Arabic speakers.

At the test level, however, minimal score differences were found between the two groups of examinees on the bottom-up and top-down bundles of items. For example, the difference in the sum of the statistically significant bottom-up bundle $\hat{\beta}_{UNI}$ values (1.87) and the sum of the statistically significant

top-down bundle $\hat{\beta}_{UNI}$ values (1.045) was 0.825. This means that at the test level, the CLBA Reading Assessment was found to favor the Mandarin-speaking examinees over equal-ability Arabic-speaking examinees by approximately one score point. This rather small difference in overall difficulty might be attributed to the fairly even balance in the number of bottom-up (18) and top-down (14) items included in Form 1, Stage II of the CLBA Reading Assessment. However, it is likely that the test would favor the Mandarin speakers to a greater extent if there were a greater proportion of bottom-up items on the test, as is the case with many reading comprehension tests (see Hill & Parry, 1992; Purpura, 1997).

Unfortunately, a content analysis of the item(s) in the three nonsignificant bundles (i.e., B4, T4, and T5) did not reveal any explanations for the departures from the predicted DIF patterns. Perhaps the presence of other unidentified construct-relevant or irrelevant dimensions is contributing to the inconsistencies in the patterns of DIF operating within these bundles. As Bolt (2002) suggested, several construct-irrelevant dimensions might work together to impact performance on items within a bundle. For example, distinct item format effects (e.g., constructed response vs. multiple choice) or passage topic effects might be regarded as additional dimensions operating within the bottom-up, top-down strategy dimensions in the CLBA Reading Assessment.

Preliminary Implications

Results from this study support Parry's (1996) conclusion that successful strategy use is a function of linguistic/cultural differences. It appears that ESL learners in these two linguistic/cultural groups have particular reading strategy strengths and weaknesses that might be related to their experiences with written language and the social process of learning to read (Parry, 1996). It might be the case that the strategies that the readers

used for coping with the orthographic features specific to their L1s were transferred to L2 reading (Akamatsu, 2003; Koda, 1988). These results also lend some support to the claim that instructional approaches might reinforce or encourage the use of bottom-up, word-level strategies over top-down, text-level strategies to overcome comprehension difficulties or vice versa (Kohn, 1992; Parry, 1993). For example, the structure of Chinese EFL textbooks and methods of teaching, such as grammar translation and structural approaches to teaching EFL, where most of the teaching and communication is in the L1, likely encourages the effective use of bottom-up strategies, whereas the exposure the Arab EFL students receive to more communicative activities likely promotes the effective use of top-down reading strategies. The results from this study also appear to support Fender's (2003) results, which imply that Arabic ESL learners have greater success in using top-down strategies than ESL learners from nonalphabetic L1 backgrounds (e.g., Chinese). It is likely that the reduction of the extent of the Arab readers' dependence on the visual stimulus in their L1 caused them to develop more effective top-down reading comprehension processes.

A number of preliminary implications for ESL reading theory and teacher, test, and curriculum development practices can also be drawn from the foregoing discussion. First, an interactive model of reading that stresses the importance of both bottom-up and top-down reading strategies appears to be a valid framework that is appropriate for modeling ESL reading comprehension in these two linguistic/cultural groups. If curriculum developers structure reading curricula using a balanced interactive approach to reading that emphasizes both bottom-up and top-down skill and strategy development, this would help learners from these two different linguistic and cultural backgrounds to be more successful readers, as it would allow them to capitalize on their strengths and receive instruction in their areas of weakness. Such an approach would reduce the current attention that many educators place on top-down processing in the construction of meaning when reading (Birch, 2002).

A further implication for practice that is evident from this research is that test developers need to be aware of the effects of L1 and culture on reading so that they can ensure that examinees' prior knowledge and cultural values and assumptions do not place examinees from specific backgrounds at a disadvantage when taking reading comprehension tests. As Hill and Parry (1992) suggested, test developers traditionally have tried to select tasks and design test items that are not offensive to any particular group of learners; however, they do not typically "use any means of evaluating how sociocultural norms of language, thought and experience are reflected in how test takers respond" (p. 455). The confirmatory DIF analysis framework employed in this study provides researchers and test developers with a method to pursue this goal.

An additional implication of this research is that if ESL teachers have a better understanding of the linguistic/cultural differences that influence successful reading strategy use, they should be able to enhance the language acquisition of adult immigrants and expedite their integration into the workplace or academia.

Limitations and Implications for Future Research

An evaluation of the relevance and representativeness of the CLBA Reading Assessment's content in relation to the construct definition used by CLBA test developers was not possible, as no information on the theory of reading or the table of specifications used to shape the assessment was available upon request. Nonetheless, assuming that an interactive theory of reading (which stresses a combination of both bottom-up and top-down reading skills) was used to guide the development of the CLBA Reading Assessment, the presence of DIF in the current study indicates that factors related to bottom-up and top-down processing affect the probability of a correct response.

One implication for future research that is evident from this research is that the verbal report sample should be increased.

Then for the DIF analyses, each CLBA item could be classified according to the most frequently used reading strategy rather than the strategy that the reading experts believed to be the most salient in answering each question. Unfortunately, however, such an increase would be even more resource- and time-intensive.

Although the results of this study suggest that DIF on the CLBA Reading Assessment appears to be associated with reading strategies that might be specific to group membership and L1 background, the analyses conducted in this study need to be replicated with different samples of Arabic- and Mandarin-speaking examinees across a variety of levels of learners. By using a confirmatory approach, researchers can continue to create a body of confirmed DIF hypotheses, which might provide further insights into the causes of DIF (Stout & Roussos, 1995). In addition, because the results of the current study were based on a limited item pool (32 items), follow-up substantive and statistical DIF studies of additional CLBA reading test forms should be conducted to determine whether similar patterns emerge for items and bundles created using the reading strategy framework. If the same statistically significant bundle differences are found in future cross-validation and generalizability studies, this would imply that the items in the bottom-up and top-down reading strategy bundles are measuring distinct secondary dimensions operating within the CLBA Reading Assessment and the same hypotheses could be tested on other ESL reading tests.

Conclusions

Theoretically, this study has drawn upon L1 and L2 reading strategy research, cognitive psychology, L2 assessment research, and psychometric research to develop a theoretical framework to test the hypothesis that some of the items included in the CLBA Reading Assessment favor certain cultural groups whose L1 orthographies differ markedly. Methodologically and analytically, this study has demonstrated the value of combining multiple sources of data and analyses (i.e., data from readers'

verbal reports, substantive item evaluation, as well as DIF analyses) to evaluate group differences on the CLBA Reading Assessment. By employing a confirmatory approach to DIF, valuable insight into the underlying causes of DIF on the CLBA Reading Assessment was gained.

Substantive analysis of the CLBA reading test items and DIF analyses based on the reading strategy framework revealed that differential skills in the application of reading strategies resulted in systematic performance differences between equal-ability Arabic- and Mandarin-speaking examinees: Items involving breaking lexical items into smaller parts, scanning for details, identifying synonyms or paraphrases, and matching key vocabulary in the text to key vocabulary in the item were found to favor the Mandarin-speaking examinees, whereas items involving skimming for gist, connecting or relating information presented in different parts of the text, and drawing an inference based on information presented in the text were found to favor the Arabic-speaking examinees.

Contrasting linguistic, cultural, and educational features of Arabic and Mandarin speakers' backgrounds were identified as potential contributors to the particular strengths and weaknesses in the successful application of Arabic- and Mandarin-speaking ESL learners' reading skills and strategies. It was likely that the Arabic- and Mandarin-speaking ESL learners' primary L1 processing strategies, which were developed through exposure to distinct languages and literary and educational practices, differentially influenced their success in using ESL reading strategies when reading and answering the CLBA reading comprehension questions. The Mandarin-speaking ESL learners appeared to be more successful at using local, detail-oriented linguistic cues and strategies, whereas the Arabic-speaking ESL learners appeared to be more successful at integrating semantic cues by relying on big-picture-oriented strategies and the global structure of text. These results have valuable implications for the theory of reading in an L2, as an interactive compensatory approach to reading that emphasizes both bottom-up and top-down reading skills and

strategies appears to be a valid framework that is appropriate for modeling ESL reading comprehension in these two linguistic/cultural groups.

Practically, this study has provided a number of preliminary suggestions for ESL teachers and language learners that ultimately could help ESL readers develop more effective reading comprehension and test-taking strategies. It appears that a balanced or interactive approach that emphasizes the importance of both bottom-up and top-down processing in the construction of meaning is appropriate for teaching reading comprehension, especially in intermediate ESL classes with students from a variety of linguistic/cultural backgrounds. In addition, this study has potentially valuable implications for test developers that might promote greater equity and fairness in reading comprehension test development practices. If the results of this study are confirmed in future research on additional reading assessments, test developers could use the information revealed about the bottom-up, top-down dimensions operating within the assessments to facilitate future item construction and the development of test specifications.

Multiple forms of evidence from additional confirmatory DIF studies have the potential to illuminate the effects of linguistic/cultural background on the validity of reading test score interpretations and inform future cross-cultural reading strategy and strategy training studies. Further studies of this nature could promote more responsible, ethical assessment practices that ensure equity in the interpretation of English language placement and proficiency reading test results and future exam and ESL course development practices.

Revised version accepted 21 March 2006

Notes

[1]Because the CLBA is a secured test, this passage and the corresponding items are not the actual CLBA Reading Assessment items that were used in the current study.

[2]Tracey Derwing and I developed the questions. Then I classified each question according to the reading strategy that was most instrumental in answering the item. The reading strategies included the following: B1, breaking lexical items into smaller parts; B2, scanning for details; B3, identifying synonyms or paraphrases; B4, matching key words to key visuals; B5, matching key vocabulary in the text to key vocabulary in the item; B6, using knowledge of grammar or punctuation; B7, using local context cues to interpret a word or phrase; T1, skimming for gist; T2, connecting or relating information.

References

Abu-Rabia, S. (1997). Reading in Arabic orthography: The effect of vowels and context on reading accuracy of poor and skilled native Arabic readers. *Reading and Writing: An Interdisciplinary Journal*, *9*, 65–78.

Abu-Rabia, S. (1999). The effect of Arabic word vowels on the reading comprehension of second- and sixth-grade native Arab children. *Journal of Psycholinguistic Research*, *28*, 93–101.

Ackerman, T., Simpson, M., & de la Torre, J. (2000, April). *A comparison of the dimensionality of TOEFL response data from different first language groups*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Akamatsu, N. (1999). The effects of first language orthographic features on word recognition processing in English as a second language. *Reading and Writing: An Interdisciplinary Journal*, *11*, 381–403.

Akamatsu, N. (2003). The effects of first language orthographic features on second language reading in text. *Language Learning*, *53*, 207–231.

Alderson, J. (1984). Reading in a foreign language: A reading problem or a language problem? In J. Alderson & A. Urquhart (Eds.), *Reading in a foreign language* (pp. 1–24). London: Longman.

Alderson, J. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Alexander, P., & Jetton, T. (2000). Learning from text: A multidimensional and developmental perspective. In M. Kamil, O. Mosenthal, P. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 285–310). Mahwah, NJ: Erlbaum.

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

Barnett, M. (1988). Reading through context: How real and perceived strategy use affects L2 comprehension. *Modern Language Journal*, *72*, 150–160.

Bedell, D., & Oxford, R. (1996). Cross-cultural comparisons of language learning strategies in the People's Republic of China and other countries. In R. Oxford (Ed.), *Language learning strategies around the world: Cross-cultural perspectives* (pp. 47–60). University of Hawaii at Manoa: Second Language Teaching and Curriculum Center.

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

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

Bolt, D. (2002, April). *Studying the potential of nuisance dimensions using bundle DIF and multidimensional IRT analyses*. Paper presented a the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Brantmeier, C. (2000). The relationship between readers' gender, passage content, comprehension and strategy use in reading Spanish as a second language. *Dissertation Abstracts International*, *61*(04), 1376. (UMI No. 9966039)

Brantmeier, C. (2003a). Beyond linguistic knowledge: Individual differences in second language reading. *Foreign Language Annals*, *36*, 33–43.

Brantmeier, C. (2003b). Does gender make a difference? Passage content and comprehension in second language reading. *Reading in a Foreign Language*, *15*(1), 1–15. Retrieved October 1, 2004, from http://nflrc.hawaii.edu/rfl/April2003/

Brown, J. (1999). The relative importance of persons, items, subtests and languages to TOEFL test variance. *Language Testing*, *16*, 217–238.

Burnaby, B., & Sun, Y. (1989). Chinese teachers' views of western language teaching: Context informs paradigms. *TESOL Quarterly*, *23*, 219–238.

Cameron, J., & Derwing, T. (2004). *Being Canadian* (2nd ed.). Saint-Laurent, Quebec: Longman.

Camilli, G., & Shepard, L. (1994). *Methods for identifying biased test items*. Newbury Park, CA: Sage.

Carrell, P. (1983). Some issues in studying the role of schemata, or background knowledge in second language comprehension. *Reading in a Foreign Language*, *1*, 81–92.

Carrell, P. (1989). Metacognitive awareness and second language reading. *Modern Language Journal*, *73*, 121–133.

Chen, H. (1992). Reading comprehension in Chinese: Some implications from character reading times. In H. Chen & O. Tzeng (Eds.), *Language processing in Chinese* (pp. 175–205). Amsterdam: Elsevier.

Chen, Z., & Henning, G. (1985). Linguistic and cultural bias in language proficiency tests. *Language Testing*, *2*, 155–163.

Coady, J. (1997). L2 vocabulary acquisition through extensive reading. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 225–237). Cambridge: Cambridge University Press.

Cohen, A. (1998). Strategies and processes in test taking and SLA. In L. Bachman & A. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 90–111). Cambridge: Cambridge University Press.

Davis, F. (1968). Research in comprehension in reading. *Reading Research Quarterly*, *3*, 499–545.

Drum, P., Calfee, R., & Cook, L. (1981). The effects of surface structure variables on performance in reading comprehension tests. *Reading Research Quarterly*, *16*, 486–514.

Duffy, G., Roehler, L., Sivan, E., Rackcliffe, G., Book, C., Meloth, M., et al. (1987). Effects of explaining the reasoning associated with using reading strategies. *Reading Research Quarterly*, *22*, 347–368.

Durgunoglu, A., & Hancin, B. (1992). An over-view of cross-language transfer in bilingual reading. In R. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 391–411). New York: Elsevier Science.

Ericsson, K., & Simon, H. (1993). *Protocol analysis: Verbal report data*. Cambridge, MA: MIT Press.

Eskey, D. (1997). Models of reading and the ESOL student: Implications and limitations. *Focus on Basics*, *1*(B), 9–11.

Fender, M. (2003). English word recognition and word integration skills of native Arabic- and Japanese-speaking learners of English as a second language. *Applied Psycholinguistics*, *24*, 289–315.

Ginther, A., & Stevens, J. (1998). Language background and ethnicity, and the internal construct validity of the Advanced Placement Spanish Language Examination. In A. Kunnan (Ed.), *Validation in language assessment* (pp. 169–194). Mahwah, NJ: Erlbaum.

Glaser, B. (1992). *Basics of grounded theory analysis: Emergence vs. forcing*. Mill Valley, CA: Sociology Press.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Green, A. (1998). *Verbal protocol analysis in language testing research: A handbook*. Cambridge: Cambridge University Press.

Gubrium, J. (1995). Taking stock. *Qualitative Health Research*, *5*, 267–269.

Harshbarger, B., Ross, T., Tafoya, S., & Via, J. (1986, March). *Dealing with multiple learning styles in the ESL classroom*. Symposium presented at the annual international meeting of Teachers of English to Speakers of Other Languages, San Francisco.

Hill, C., & Parry, K. (1989). Autonomous and pragmatic models of literacy: Reading assessment in adult education. *Linguistics and Education*, *1*, 233–283.

Hill, C., & Parry, K. (1992). The test at the gate: Models of literacy in reading assessment. *TESOL Quarterly*, *26*, 433–461.

Hosenfeld, C. (1977). A preliminary investigation of the reading strategies of successful and non-successful language learners. *System*, *5*, 110–123.

Katz, L., & Frost, R. (1992). Reading in different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 67–84). Amsterdam: Elsevier.

Kharma, N. (1998). EFL and community needs. *International Review of Applied Linguistics in Language Teaching*, *36*, 49–69.

Koda, K. (1988). Cognitive processes in second language reading: Transfer of L1 reading skills and strategies. *Second Language Research*, *4*, 133–156.

Kohn, J. (1992). Literacy strategies for Chinese university learners. In F. Dubin & N. Kuhlman (Eds.), *Cross-cultural literacy* (pp. 113–125). Englewood Cliffs, NJ: Regents.

Kunnan, A. (1994). Modelling relationships among some test-taker characteristics and performance on EFL tests: An approach to construct validation. *Language Testing*, *11*, 225–252.

Levine, A., Reves, T., & Leaver, B. (1996). Relationship between language learning strategies and Israeli versus Russian cultural-educational factors. In R. Oxford (Ed.), *Language learning strategies around the world: Cross-cultural perspectives* (pp. 35–46). University of Hawaii at Manoa: Second Language Teaching and Curriculum Center.

Liontas, J. (1999). Developing a pragmatic methodology of idiomaticity: The comprehension and interpretation of SL vivid phrasal idioms during reading. *Dissertation Abstracts International*, *60*(09), 3346. (UMI No. 9946784)

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, *13*, 241–256.

Munby, J. (1978). *Communicative syllabus design*. Cambridge: Cambridge University Press.

O'Malley, J., & Chamot, A. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.

Oxford, R. (1990). *Language learning strategies*. New York: Newbury House.

Parry, K. (1993). The social construction of reading strategies: New directions for research. *Journal of Research in Reading*, *16*, 148–158.

Parry, K. (1996). Culture, literacy and L2 reading. *TESOL Quarterly*, *30*, 665–692.

Penner, J. (1995). Change and conflict: Introduction of the communicative approach in China. *TESL Canada Journal*, *12*(2), 1–17.

Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, *20*, 26–56.

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

Pritchard, R. (1990). The effects of cultural schemata on reading processing strategies. *Reading Research Quarterly*, *25*, 273–295.

Purpura, J. (1997). An analysis of the relationships between test takers' cognitive and metacognitive strategy use and second language test performance. *Language Learning*, *47*, 289–325.

Reid, J. (1995). *Learning styles in the EFL/ESL classroom*. Boston: Heinle & Heinle.

Rennie, D. (1984, May). *Clients' tape-assisted recall of psychotherapy: A qualitative analysis*. Paper presented at the annual meeting of the Canadian Psychological Association, Ottawa, Ontario.

Robb, L. (1996). *Reading strategies that work: Teaching your students to become better readers*. New York: Scholastic.

Roussos, L., & Stout, W. (1996). A multidimensionality-based DIF analysis paradigm. *Applied Psychological Measurement*, *20*, 355–371.

Ryan, K., & Bachman, L. (1992). Differential item functioning on two tests of EFL proficiency. *Language Testing*, *9*, 12–29.

Sasaki, M. (1991). A comparison of two methods for detecting differential item functioning in an ESL placement test. *Language Testing*, *8*, 95–111.

Scheuneman, J. (1979). A new method for assessing bias in test items. *Journal of Educational Measurement*, *16*, 143–152.

Schueller, J. (2004). Gender and foreign language reading comprehension: The effects of strategy training. *Southern Journal of Linguistics*, *27*(1), 45–65.

Shealy, R., & Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DIF as well as item bias/DIF. *Psychometrika*, *58*, 159–194.

Shepard, L., Camilli, G., & Averill, M. (1981). Comparison of six procedures for detecting test item bias using both internal and external ability criteria. *Journal of Educational Statistics*, *6*, 317–375.

Standards for Educational and Psychological Testing. (1999). Washington, DC: American Educational Research Association, American Psychological Association, & National Council on Measurement in Education.

Storey, P. (1997). Examining the test-taking process: A cognitive perspective on the discourse cloze test. *Language Testing*, *14*, 214–231.

Stout, W., & Roussos, L. (1995). *SIBTEST manual*. Chicago: University of Illinois at Urbana-Champaign, Department of Statistics, Statistical Laboratory for Educational and Psychological Measurement.

Stout, W., & Roussos, L. (1999). *Dimensionality-based DIF/DBF package* [Computer software]. Chicago: William Stout Institute for Measurement, University of Illinois.

Weinstein, C., & Mayer, R. (1986). The teaching of learning strategies. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 315–327). New York: Macmillan.

Willing, K. (1988). *Learning styles in adult migration education*. Adelaide: National Curriculum Resource Centre.

Young, D., & Oxford, R. (1997). A gender-related analysis of strategies used to process input in the native language and a foreign language. *Applied Language Learning*, *8*, 43–73.

Appendix

Instructions:

Please read the passage below and answer the following questions.

Environmental Issues

The environment has become an important issue in Canada and other parts of the world. Many scientists say that if we don't change our way of living, we will destroy the world. What happens in one country can affect the rest of the world.

One issue that has received a great deal of attention is global warming. Many scientists believe that the whole earth is becoming a giant greenhouse. The earth is surrounded by a blanket of gases called the atmosphere. The gases act like the glass in a greenhouse, trapping the heat from the sun. Recently, there has been a striking increase in certain gases, especially carbon dioxide, methane, and chlorofluorocarbons; these gases trap heat. Consequently the average temperature of the earth is rising, a trend called global warming.

Global warming is caused by the burning of fuels such as oil, coal, wood and gas; deforestation; vehicle exhaust;

fertilizers; rotting garbage; and cattle digesting food. In fact, most things that consume energy contribute to the problem (e.g., air conditioning, heating, driving, and manufacturing). Canada has one of the worst records of the industrialized nations for producing greenhouse gases (only the United States and Luxembourg have worse records than Canada).

Global warming results in frightening consequences to the climate. A hotter earth means that ice caps in the polar region will melt, causing oceans to rise. Many islands will disappear under the water and coastal areas will be flooded. Studies estimate that 35% of Bangladesh will be under water by the year 2100. Many plants, fish and animals will be unable to survive the warmer temperatures. Some parts of the world will get less rain and crops will suffer.

In the summer of 2001, for example, the Prairies suffered the driest summer ever; many farmers had no crops and could not feed their animals.

The drought situation is even worse in Africa, where more and more land becomes desert every year. All countries contribute to the global warming problem, but the industrialized nations are the worst offenders.

In 1997, Canada along with another 160 countries, met in Kyoto, Japan to discuss how to reduce greenhouse gases around the world. The countries set targets for lower production of gases; the agreement to achieve these targets is called the Kyoto protocol. The Kyoto protocol was signed by Canada in 2002. Many politicians and business people are concerned that agreements such as the Kyoto protocol will result in job loss and a poor economy[1].

Note. From *Being Canadian* (p. 94), by J. Cameron & T. Derwing, 2004 (2nd ed.), Saint-Laurent, Quebec: Longman.

Questions: (Please circle the letter of the best answer).	Strategy[2]
1. The word deforestation in paragraph three means (a) deacidifying forests. (b) replanting forests. (c) clearing forests. (d) preserving forests.	B1
2. The word striking in the second paragraph means (a) slight. (b) painful. (c) significant. (d) predictable.	B3
3. Which of the following pictures represents a fuel that was specified in the passage? (a)  (b)  (c)  (d) 	B4
4. The primary cause of global warming is (a) fish farming. (b) heat from the sun. (c) the burning of fuels. (d) the melting ice caps.	B5
5. Which of the following words is not similar in meaning to the word *consequence*? (a) outcome (b) result (c) effect (d) affect	B6
6. The word drought in paragraph six means (a) dry (b) rainy (c) global (d) political	B7
7. Which of the following would be the best subtitle for the article? (a) The Elimination of Chlorofluorocarbons (b) The Production of Greenhouse Gases (c) The Signing of the Kyoto Protocol (d) The Effects of Global Warming	T1
8. What actions could be taken to slow down global warming?	T2

(a) intensified fertilizer use
(b) reduced recycling
(c) additional cattle
(d) reduced driving

9. The reader cannot conclude that T3
(a) the polar ice caps are melting.
(b) Luxembourg consumes a lot of energy.
(c) Canada has met the Kyoto Protocol target.
(d) the drought conditions in Africa are very bad.

10. Which of the following is a consequence of global warming? T5
(a) increasing energy consumption
(b) increasing the death of plants
(c) increasing world population
(d) increasing greenhouse gases

Please fill in the following blanks with the best answer.

11. Canada signed the Kyoto protocol in _____. B2

12. Bangladesh is a_____. T4