INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

University of Alberta

ENGLISH READING STRATEGIES: DIFFERENCES IN ARABIC AND MANDARIN SPEAKER PERFORMANCE ON THE CLBA READING ASSESSMENT

by

Marilyn Lorraine Abbott 

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in

Measurement, Evaluation, and Cognition

Department of Educational Psychology

Edmonton, Alberta

Spring 2005

Dedication

I lovingly dedicate this work in memory of my dear father Gerald Hamilton Abbott, who instilled in me the values of lifelong learning, courage, hard work, and perseverance, yet never forgot to remind me of the real priorities in life. I will forever be grateful for his example and inspiration.

Abstract

This study was undertaken to test the hypothesis that some of the items included in Form 1, Stage II of the Canadian Language Benchmarks Assessment (CLBA) reading subtest favour certain cultural groups whose first language orthographies differ markedly. It was posited that Mandarin speakers, who have a tendency to use bottom-up, local reading strategies, would perform better on particular questions than Arabic speakers, who have a tendency to use top-down, global reading strategies and vice versa. In Part 1 of the study, verbal report data were collected from Arabic and Mandarin speaking intermediate ESL learners to identify, clarify, and elaborate on the reading strategies involved in carrying out the CLBA reading comprehension tasks.

In Part 2 of the study, two samples of examinees were drawn from previously administered CLBA Form 1, Stage II Reading Assessments. One sample consisted of 250 Arabic speaking immigrants, and the other consisted of 250 Mandarin speaking immigrants matched for education and language level on time of arrival. Three ESL reading experts classified each of the 32 CLBA reading items into one of the seven bottom-up or five top-down reading strategy categories that had emerged from the data in Part 1 of this study. Differential bundle functioning (DBF) analyses were then conducted to determine whether groups of CLBA items classified according to the bottom-up, top-down organizing principle functioned differentially for equal ability Arabic and Mandarin ESL learners. SIBTEST analyses revealed systematic group differences in four of the bottom-up strategy categories and three of the top-down categories. Items involving breaking words 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 favour the Mandarin speaking examinees. 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 favour the Arabic speaking examinees. These results provide (a) evidence for the validity of the bottom-up, top-down reading strategy framework and (b) a substantive method for interpreting group differences on the CLBA Reading Assessment.

Acknowledgements

One does not complete a project like this alone. Working on this dissertation has been a rewarding experience thanks to the efforts of so many people.

I am very appreciative of my committee for their assistance with this study. First, I would like to thank Dr. Tracey Derwing, my supervisior, for her unwavering support, compassion, counsel, and exemplary scholarship. Dr. Derwing was one of the two primary sources of inspiration for this project. Her sense of determination and persistence was fundamental to getting this study off the ground and nurturing it through to completion. She spent too many hours to count helping me gain permission to conduct this research. I am sure that I could never express to her the depth of my appreciation for all she has done to make this project a reality.

Second, I would like to thank Dr. Mark Gierl who was the second source of inspiration for this project. The methodological and analytical foundations for this research were developed in Dr. Gierl's graduate courses. His extensive knowledge of differential item and bundle functioning and the outstanding research he has conducted in these areas have impacted my research in many positive ways. I am a far more knowing researcher for having studied under him.

Third, I would like to thank Dr. W. Todd Rogers for his continued support and attention to detail, and the key role he played in helping me gain permission to study the CLBA Reading Assessment. I am also grateful for the intellectual stimulation provided by Dr. Rogers as he raised my mind to levels of thinking that I had never imagined possible.

I would like to express my heartfelt thanks to my external examiner, Dr. Neil Anderson, for his thoughtfulness and the insightful comments that he made on my work. The expert contributions that this master practitioner has made to the ESL reading strategies literature have had an enormous impact on my research.

I would like to thank the rest of my committee, Dr. Judy Cameron and Dr. Carol Leroy, for their helpful comments and suggestions on various aspects of this dissertation.

I give thanks for the wise and gentle counsel of Dr. Marian Rossiter. Her friendship, support, and encouragement were instrumental in the attainment of this degree.

I am grateful to my classmates for their comradeship. I am particularly indebted to Jie Lin who spent countless hours assisting me with my data collection. Jie also provided encouragement and insightful feedback as my study took shape.

I would like to thank the Arabic and Mandarin speaking students for volunteering their time and effort for this study. I am also grateful for the assistance provided by Jaye Fredrickson, Anna DeLuca, Elena Chernaeva, and all of the ESL instructors at NorQuest College who encouraged their students to participate in this study.

I also wish to thank the following people for their approval and support of this research project: Grazyna Pawlikowska-Smith, author of the Canadian Language Benchmarks; Carolyn Cohen of the Centre for Language Training and Assessment; Pauline McNaughton of the Centre for Canadian Language Benchmarks; Dr. Philip Nagy of the University of Toronto; Rob Vineberg, Randy Gurlock, Gayle Taylor, and Robert Ferguson of Citizenship and Immigration Canada; Carolyn Dieleman of Alberta Advanced Education; AnaMaria Fantino, Annette Kreider, Heather Plaizier, and Bason

Hoang of the Language Assessment Referral and Counselling Centre in Edmonton; and Vivian Chan and Salena Ukueku of the Immigrant Language Vocational Assessment - Referral Centre in Calgary. This study would never have been accomplished without their support.

I would like to mention that funding for this research has been provided by an Isaac Walton Killam Memorial Scholarship, a Social Sciences and Humanities Research Council 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. I am very appreciative of this financial support.

I would also like to thank my family and friends for always being there when I needed them. My special thanks go to my mother for her everlasting concern and understanding.

Finally, I am forever grateful to Gary Latawiec for all of his assistance with this project, and for his love and continued support throughout the entire process. He was the one who encouraged and cajoled me to pursue doctoral studies. His belief in me helped me to persevere. Thus, I share this accomplishment with him.

Table of Contents

CHAPTER I: INTRODUCTION 1

Research Questions 4

CHAPTER II: LITERATURE REVIEW 6

Overview 6

Reading Comprehension 6

Defining Reading Comprehension 6

Knowledge that Affects Comprehension 9

Strategic Processing 9

Defining Reading Comprehension Strategies 12

Reading Strategy Research 14

Methods and Issues Concerning the Study of Reading Strategies 14

Second Language Reading Strategies Inferred from Verbal Reports and Recalls 14

Reading Strategy Questionnaire Studies 23

Strategy Use and ESL/EFL Test Performance 28

Effects of Culture on Reading Strategy Use 34

Difficulties in Comparing Reading Strategy Studies 37

Second Language Reading Strategy Classification Schemes 38

Linguistic Cues 38

Examples of Bottom-up Local and Top-down Global Reading Strategies 40

A Cultural-Historic Approach to Strategy Research 41

The Influence of Culture, Education, and Language on Strategy Use 43

The Effects of Linguistic Differences on Strategy Use 45

Reading Assessment 47

Differential Item Functioning (DIF) 48

DIF on ESL/EFL Proficiency and Placement Tests 49

A Confirmatory Approach to DIF 54

The Simultaneous Item Bias Test 54

Literature Summary 57

CHAPTER III: METHOD 60

Overview 60

Method – Part 1 Verbal Report 61
Participants 61
Instruments 62
Procedure 64
Data Analysis 67
Method – Part 2 Differential Item and Bundle Functioning 69
Sample 69
Procedure 70
Data Analysis 72

CHAPTER IV: RESULTS 73
Part 1 - Verbal Report Results 73
Background Questionnaire 73
Canadian Language Benchmarks Assessment - Reading Assessment 74
Topic Familiarity, and Perceived Passage and Item Comprehension 74
Reading Ability, Preferences, and Strategies Questionnaire 77
Interrater Agreement 84
Coding Schema 85
Frequencies of Bottom-up and Top-down Reading Strategies 87
Part 2 - Differential Item and Bundle Functioning Results 92
Sample 92
Item Coding 94
Psychometric Characteristics of the CLBA Reading Assessment 101
Single-Item SIBTEST Results 101
Differential Bundle Functioning Results 104

CHAPTER V: DISCUSSION AND CONCLUSIONS 108
Summary of Research Questions and Hypotheses 108
Discussion 109
Research Question 1 109
Research Question 2 113
Research Question 3 115
Supplementary Data 120
Limitations and Implications for Future Research 127

Verbal Report and Supplementary Data Limitations and Implications 127
Differential Item and Bundle Functioning Limitations and Implications 129
Conclusions 133
References 137
Appendix A: Letter of Invitation to Participate in the Study 155
Appendix B: Language Background Questionnaire 156
Appendix C: Reading Ability, Preferences, and Strategies Questionnaire 157
Appendix D: Confidentiality Agreement for Bilingual Interpreters 159
Appendix E: ESL Participant Consent Form 160
Appendix F: Verbal Report Practice Instructions 161
Appendix G: Confidentiality Agreement for Researcher and Referral Centres 163
Appendix H: Confidentiality Agreement for Assistant and Referral Centres 164
Appendix I: ESL Expert Consent Form 165
Appendix J: CLBA Reading Assessment Coding Sheet 166

List of Tables

Table 1: Demographic Information for Verbal Report Participants 74

Table 2: Distributions of CLBA Topic Familiarity by Language Group 75

Table 3: Distributions of CLBA Passage Comprehension by Language Group 76

Table 4: Distributions of CLBA Item Comprehension by Language Group 78

Table 5: Perceived L1 and L2 Reading Ability and Preferences for Learning English 79

Table 6: Self-reported Bottom-up Reading Strategies by Language Group 81

Table 7: Self-reported Top-down Reading Strategies by Language Group 83

Table 8: Reading Strategies Used When Answering the CLBA Reading Items 86

Table 9: Bottom-up, Top-down Strategy Frequencies and Proportion Scores for Each Participant and Language Group 88

Table 10: CLBA Reading Scores Compared with Proportions of Strategies Used 90

Table 11: Demographic Information from Part 2 of the Study 93

Table 12: Comparing Demographic Information from Part 1 and Part 2 of the Study ... 94

Table 13: Interrater Agreement for Three Expert Item Reviewers 96

Table 14: Distribution of CLBA Reading Items Across the Bottom-up and Top-down Reading Strategy Categories 98

Table 15: All Salient Strategies Identified as Useful When Answering Each Item 100

Table 16: Descriptive Statistics for the CLBA Reading Assessment 102

Table 17: CLBA Single-item SIBTEST Results 103

Table 18: Differential Bundle Functioning Results 107

List of Figures

Figure 1. Basic Elements of the Human Information-Processing System 7

Figure 2. Differential Item Functioning Results 105

CHAPTER I: INTRODUCTION

The Canadian federal government provides language training to immigrants who have limited or no proficiency in an official language on arrival. Many newcomers, however, are unable to access more than the federally funded maximum of 1500 hours of instructional support. Consequently, it is necessary that immigrants' language levels be accurately assessed so they can be placed in the most appropriate levels of instruction. Otherwise their time and the federal support they receive will be wasted.

It is also crucial to ensure that placement tests provide equal opportunities for all immigrants to demonstrate what they know about "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 may unfairly favour 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 may be easier for individuals of that group. Fair, equitable assessment is tailored to the individual learner's instruction context and background including his or her prior knowledge, cultural experience, language proficiency, cognitive style, and interests (*Principles for Fair Student Assessment Practices for Education in Canada*, 1993). Therefore, both substantive and statistical research devoted to examining and promoting accuracy and fairness when developing and using assessment tools such as the Canadian Language Benchmarks Assessment (CLBA) is essential.

Since its inception in 1996, the 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 post-secondary institutions. For example, CLBA results are recognized by the Southern Alberta Institute of Technology. This usage has moved the CLBA into the realm of high-stakes testing.

To date, the extent to which the CLBA reading items may favour 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 the current study was to develop a theoretical reading strategy framework to test the hypothesis that some of the items included in the CLBA Reading Assessment favour Arabic speaking examinees over Mandarin speaking examinees and vice versa. For example, it was posited that Mandarin speakers, who have a tendency to use bottom-up, local, word-level reading strategies, would perform better on particular questions than Arabic speakers, who tend to use top-down, global, text-level reading strategies. Differential item and bundle functioning analyses were conducted to determine whether single items and groups of CLBA reading items classified according to the theoretical bottom-up, top-down framework, functioned differentially for Arabic and Mandarin first language immigrant groups. 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.

While a plethora of questionnaire research results indicate that cultural background affects second language 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 second language reading strategies interact with first language 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, whereas top-down strategies are conceptually- or hypothesis-driven (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 cross-linguistic 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 may 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 first language research (which currently emphasizes the importance of visual processing), second language reading researchers (e.g., Brantmeier, 2000, 2003a, 2003b; Coady, 1997; Eskey, 1997; Liontas, 1999; Parry, 1996; Schueller, 2000, in press; Storey, 1997; Young & Oxford, 1997) currently rely largely on bottom-up, top-down models of reading comprehension. Therefore, the analyses conducted in this study were also based upon a bottom-up, top-down approach to reading.

Four characteristics distinguish this study from previous research. Although most of the previous studies of second language learners' reading strategies that argue for a bottom-up/local and top-down/global classification schema were based on rather small sample sizes, the current study tested this framework using data collected from a much larger sample (i.e., 250 Arabic and 250 Mandarin speakers). In addition, as the differential item functioning (DIF) literature indicates, DIF research on ESL placement and proficiency tests has been rather sparse and a confirmatory approach to the study of DIF in such exams has never been employed. Furthermore, no DIF research has ever been conducted which has compared the performance of equal ability Mandarin and Arabic speaking examinees on an ESL placement or proficiency test. Finally, no such research has involved recently arrived immigrants studying ESL outside of a university context.

Research Questions

Three research questions were investigated in this study, which was conducted in two parts. Question 1 was addressed by the verbal report data collected in Part 1 of the

study. Then the reading strategies identified in the verbal report data were used to fine-tune the theoretical framework that was used to classify the CLBA items in Part 2 of the study so differential item and bundle functioning analyses could be used to address questions 2 and 3.

1. What are the bottom-up and top-down reading strategies that intermediate proficiency Arabic and Mandarin speaking ESL learners employ when reading and answering the CLBA reading items?
2. Is there evidence for differential item performance for Arabic and Mandarin speaking examinees on the CLBA Reading Assessment?
3. If so, is the source of differential performance related to differences in reading strategy use?

The development of a theoretical framework for evaluating whether the CLBA Reading Assessment produces comparable results for Arabic and Mandarin first language speakers requires (a) an examination of the literature pertaining to reading comprehension, reading strategies, reading assessment, and the effects of culture, education, and first language on the development and use of ESL reading strategies; and (b) the consideration of key psychometric concepts such as differential item functioning (DIF) and differential bundle functioning (DBF). These concepts and related literature are discussed and critically reviewed in the next chapter.

CHAPTER II: LITERATURE REVIEW

Overview

The literature covered in this chapter has specific relevance to the theoretical reading strategy framework that was developed in the current study to test the hypothesis that some of the items included in the CLBA Reading Assessment function differentially for equal ability Arabic and Mandarin speaking examinees. This chapter is organized in five main sections. In the first section, the literature relevant to reading comprehension, which is what the CLBA Reading Assessment was designed to measure, is reviewed. The second section contains critical reviews of research pertaining to English as a second language (ESL) and English as a foreign language (EFL) reading strategies and the effects of culture, education, and first language on the development and use of reading strategies. A brief discussion of the purposes and concerns of reading assessment specialists in the third section leads into a discussion of the psychometric literature of specific relevance to identifying differential item and bundle functioning on the CLBA Reading Assessment in the fourth section. In the final section, a summary of the main conclusions drawn from the literature is presented.

Reading Comprehension

Defining Reading Comprehension

Reading comprehension is often defined and assessed in two ways: in terms of decoding, which refers to an examinee's ability to recognize words and recall their meanings; and in terms of learning that results from the examinee interacting with and making inferences from the text (Kinsche, 1986). According to Gagné, Yekovich, and Yekovich's (1993) model of information processing (see Figure 1), human learning is a

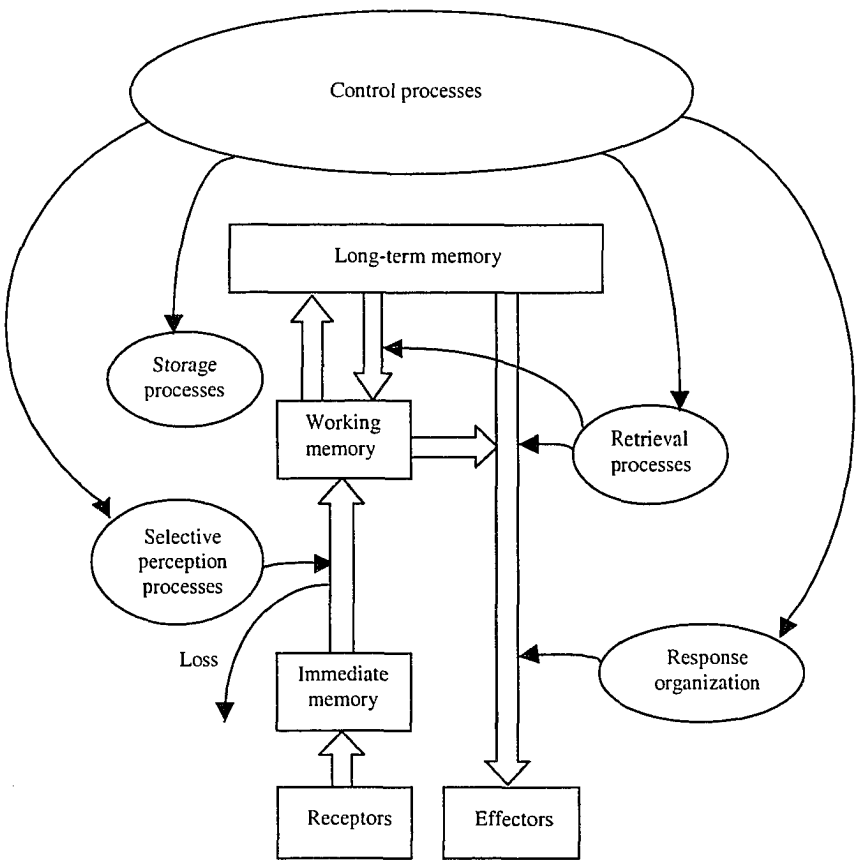


Figure 1. Basic elements of the human information-processing system (Gagné, Yekovich, & Yekovich, 1993, p. 40).

process where information is received by receptors in sensory form and registered in immediate memory through nerve impulses sent to the central nervous system. Only a small portion of information in immediate memory is selected from other input for processing. The selected information is held in working or short-term memory long enough to be transformed into meaningful symbols that are retrieved from long-term memory and related to previous knowledge structures or elaborated upon and turned into grammatical forms which are sent out to the environment through effectors (e.g., writing, talking, typing) when needed. This chain of events is guided and monitored by control processes.

Reading comprehension defined in terms of decoding and learning involves a complex set of interacting processes that individuals use to construct meaning from written symbols. Although different models of reading comprehension focus on different component processes, it is generally accepted that reading has three overarching meaning acquisition processes. Rayner and Pollatsek (1989) refer to these three subcomponents as orthographic, phonological, and metacognitive processes. Orthographic processing involves translating the text into a visual code and activating the meaning of the code. Phonological processing refers to translating the text into a sound code and activating its meaning. Metacognitive processes (defined as control processes in Gagné et al.'s (1993) model) guide attention and aid in the learner's selection of appropriate knowledge of real-world events and linguistic features that help the reader construct meaning. Each of these processes requires different sets of knowledge and strategies that are used to activate and construct meaning.

Knowledge that Affects Comprehension

According to Widdowson (1983), comprehension is dependent on two main sources of knowledge: schematic and systemic knowledge. Schematic knowledge refers to background knowledge of factual and sociocultural information and knowledge of how information is used in discourse. Systemic knowledge refers to knowledge of syntax, semantics, morphology, and phonology. Skehan (1998) suggests that second or foreign language learners are likely to exploit what they know in terms of relevant schematic knowledge to overcome their linguistic limitations (i.e., lack of systemic knowledge). However, it is likely that learners also utilize their knowledge of cognitive strategies to overcome their lack of linguistic knowledge. In this sense, cognitive strategies may be defined as the mental procedures or operations that learners use to compensate for their lack of linguistic knowledge.

Strategic Processing

Typically, cognitive psychologists identify two types of knowledge that are associated with strategic processing: declarative knowledge about facts, theories, events, and objects, and procedural knowledge about how to do things. In addition to these two types of knowledge, Paris, Lipson, and Wixon (1983) identify a third category, namely, conditional knowledge of when, where, and why certain strategies are valuable. In terms of strategy use, declarative knowledge refers to a learner's explicit knowledge of what a strategy is, procedural knowledge relates to how to use the strategy, and conditional knowledge includes when and where to use the strategy and why it is important to use it in various circumstances.

Anderson (1983) identifies three key metacognitive processes – planning, selective attention, and monitoring – that are involved in directing the course and success of language comprehension through the selection and application of cognitive strategies. In this view, metacognitive strategies refer to the mental processes of assessing the situation and then selecting, enacting, and monitoring a plan to attain a goal. In the case of assessment, when solving a reading comprehension task, readers must think and act strategically by processing a variety of thoughts and plans to complete the task. The process begins with the goal of completing the reading comprehension task correctly. Consequently, strategic processing in reading assessment may be viewed as a dynamic, goal-directed process, which moves from the reader's perception of the task, to the activation of strategic competence, to strategy selection and implementation, and, lastly, to the evaluation of the result.

Strategic competence is one of the four components of communicative language ability proposed by Canale and Swain (1980). The other three components include linguistic, sociolinguistic, and discourse competence. While Canale and Swain view strategic competence as largely limited to compensatory communication strategies for dealing with communication breakdowns or for enhancing communication, Bachman (1990) and Bachman and Palmer (1996) expand the role of strategic competence to encompass a set of metacognitive strategies, namely, goal setting, assessment, and planning. Although Anderson (1983), Gagné et al. (1993), and Bachman and Palmer (1996) identify slightly different metacognitive/control processes, they concur that these strategic processes oversee and manage the selection, deployment, and evaluation of cognitive strategies.

When readers selectively attend to certain cues or features in the language input, the cues they attend to may be assumed to affect the plans that they make for completing a task. For example, when answering reading assessment questions, readers may consciously plan (at least to some degree) to focus on local word-based or global text-based cues to aid in their comprehension of the text. If the readers' past experiences and training have primarily taught them to rely on linguistic segments and features of the input when decoding text, then they will likely have a tendency to rely on bottom-up, local reading comprehension strategies. Alternately, if the readers' past experiences and training have mainly stressed the importance of experiencing overall patterns of language in meaningful contexts before making sense of the linguistic parts and forms, then they will more likely rely on top-down, global strategies.

According to schema theory (Bartlett, 1932; Rumelhart, 1980), individuals understand new information only when they connect it to something they already know. With respect to reading, features of the input may be interpreted through the best-fitting bottom-up or top-down processes. Individuals who tend to make use of top-down processing strategies utilize contextual information and their existing schematic knowledge of real-life situations and discourse organization to make meaning by anticipating and predicting what will come next in the text and making inferences about portions of the text that are not fully understood (Howard, 1985). Hence top-down strategies are meaning-directed as they rely on schematic knowledge and scripts for constructing meaning. Scripts are special schemata that consist of situation-specific knowledge about the goals, participants, procedures, and organization of real-life events in time (Schank & Abelson, 1977).

In contrast, individuals who pay closer attention to linguistic forms by analyzing each word or grammatical feature and using this information to build up meaning are relying on bottom-up processing strategies (Howard, 1985). Thus bottom-up strategies are reproduction or input-oriented as they focus on using linguistic knowledge to interpret text on an element-by-element basis. Although recent cognitive theory suggests that effective comprehension requires a combination of these two processes, the weight or amount of attention given to each of the processes will depend upon text familiarity, task demands, and individual difference variables (Stanovich, 1980, 1986, 2000). This implies that readers may use top-down strategies to compensate for poorly developed bottom-up strategies and vice versa.

Defining Reading Comprehension Strategies

Research into second language reading comprehension strategies has proved to be a complex endeavour 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 behaviours that language learners employ in second language acquisition, second language use, or second language testing situations (Alderson, 1984; Carrell, 1989; Cohen, 1998; Hosenfeld, 1977; O'Malley & Chamot, 1990; Oxford, 1990; Purpura, 1997). According to Cohen (1998), 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 may be defined as the mental operations or comprehension processes that readers select and apply in order to make sense of what they read. Since

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, subskills or microskills, others (Alexander & Jetton, 2000; Carrell, 1989; Duffy, Roehler, Sivan, Rackcliffe, Book, Meloth, Vavrus, Wesselman, Putnam, & Bassiri, 1987; Robb, 1996; Routman, 1994) refer to these behaviours 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. Thus, when a skill is used intentionally to achieve a goal, it becomes a strategy (Paris, Wasik, & Turner, 1991). This assumption was adopted in the current study. Routman (1994) summarizes this distinction between reading skills and strategies in that skills are only useful to the extent that they can be applied strategically in an authentic reading context (p. 298).

The awareness of the mental processes involved in using reading skills as strategies is referred to as metacognitive awareness (Duffy et al., 1987). When language learners consciously use strategies to process input, they are using strategies that are metacognitive in nature. Flavell (1976) describes metacognition as

> one's knowledge concerning one's own cognitive processes ... the active monitoring and consequent regulation and orchestration of these processes in

relation to the cognitive objects or data on which they bear, usually in the service of some concrete goal or objective. (p. 232)

Reading Strategy Research

Methods and Issues Concerning the Study of Reading Strategies

A variety of research methods and reading tasks have been used to explore reading strategies. Researchers have compiled a considerable number of strategies elicited by concurrent and retrospective verbal reports, self-report questionnaires, and written recalls using a range of tasks that vary in terms of topic, syntactic complexity, vocabulary, cohesion, and text type. Much reading strategy research has been conducted by interviewing secondary and university students about their strategies while they read a single experimenter-selected text in a laboratory setting at a single point in time. Reading strategy use has also been examined in settings where grade school students were asked to read strategically (Brown & Pressley, 1994; Hart & Speece, 1998; Magliano, Graesser & Trabasso, 1999; Palincsar & Brown, 1984). However, few studies of adult beginner or intermediate language learners have been carried out in non-postsecondary education settings such as adult education programs, community-based programs, and workplace literacy programs (Hacker, 1998). Consequently, the existing knowledge of strategic processing during reading may be uncharacteristic of typical reading and may have little transfer to normal reading in real-world contexts (Hacker, 1998) or to other non-college or non-university language learning and language use situations.

Second Language Reading Strategies Inferred from Verbal Reports and Written Recalls

Examinee responses collected from verbal reports and written recalls have provided some insights into reading processes and a better understanding of the strategic

processes underlying test-taking behaviour. In a qualitative study of reading strategy use, Hosenfeld (1977) compared the reading strategies reported by 20 successful and 20 unsuccessful native English speaking high school students enrolled in second year French, Spanish, or German classes. The 40 students were selected on the basis of high and low English reading test scores from a sample of 210 students who wrote the MLA-Cooperative Test of Reading Proficiency. After training in the think-aloud procedures, the students were asked to verbalize their thought processes while reading a text in the foreign language that they were studying in school. Hosenfeld's protocol analyses revealed that the successful readers kept the meaning of the passage in mind while reading, read in longer chunks, skipped nonessential words, used context as clues to unknown words, and had positive self-concepts as readers. In contrast, unsuccessful readers frequently lost the meaning of sentences while decoding, read on a word-by-word basis or in short phrases, infrequently skipped words, looked up unknown words, and had negative self-concepts as readers. Hosenfeld was the first to divide reading strategies into binary categories that reflected "main-meaning line" and "word-solving strategies" which roughly correspond to the broadly semantic top-down and narrowly linguistic bottom-up classifications outlined above on strategic processing. Unfortunately, however, Hosenfeld did not report the frequency of strategy use in each of the categories.

In another verbal report study, Block (1986) examined the reading comprehension strategies of three native and six non-native (3 Chinese, 3 Spanish) speakers of English who had failed a college-entrance reading proficiency test and as a result were enrolled in remedial native speaker and ESL reading classes. All participants achieved similarly low scores on the reading test. The six non-native participants were selected on the basis of

their high oral fluency in English. All participants were given two sample reading passages and were told to report what they were thinking and what they were doing to understand the text after reading each sentence silently to themselves. Once they were comfortable with the method, the participants were asked to provide verbal reports after each sentence while reading two passages taken from an introductory psychology text. After completing each verbal report, the students were asked to recall and report everything they remembered about the passage. Then they were asked to answer 20 multiple-choice reading comprehension questions. However, this was not a good measure of reading comprehension as the students were allowed to refer back to the text after they had completed the verbal recalls. It was possible that the acts of reporting verbally and then recalling the passages influenced the way they processed the text when answering the questions.

Like Hosenfeld (1977), Block (1986) classified the strategies identified in the verbal report passage data into two categories: general strategies and local strategies. General strategies involved developing and monitoring comprehension (e.g., anticipating content, recognizing text structure, integrating information, questioning information, interpreting, using general knowledge and associations, commenting on behaviour or process, monitoring comprehension, correcting behaviour, and reacting to the text). Local strategies involved understanding specific linguistic units (e.g., paraphrasing, rereading, questioning meaning of a clause or sentence, questioning meaning of a word, and solving a vocabulary problem). An analysis of the patterns of strategy use did not reveal any differences between (a) native and non-native English speaker strategy use, or (b) Chinese and Spanish ESL learner strategy use. However, Block found four characteristics

that distinguished the more successful from the less successful readers. The students who remembered more about the passages and scored higher on the multiple-choice comprehension questions integrated new information with old information, recognized aspects of the text structure, used their background knowledge, monitored their understanding, and responded in the extensive mode, which suggested that they focused on understanding the author's ideas without relating the text to themselves affectively or personally. In contrast, the less successful readers seldom integrated propositions, rarely recognized text structure, relied on personal experiences, and responded in a reflexive mode, which suggested that they focused on their own thoughts and feelings rather than on the information stated in the text.

Sarig (1987) examined the relationship between the first language (L1) and second language (L2) reading strategies of 10 female native Hebrew speakers who were studying EFL in high school. Participants were selected on the basis of teacher evaluations and performance on an English proficiency test to represent learners of low, intermediate, and high English proficiency levels. While reading academic texts in the L1 and L2, participants were asked to report what they were thinking. Four general types of strategies were identified in the think aloud data: technical aid, clarification and simplification, coherence detection, and monitoring strategies. Technical aid strategies included skimming, scanning, skipping words, marking and writing key elements in the text, and using the glossary. Clarification and simplification strategies included syntactic and ideational simplification, decoding meanings of words and groups of words using synonyms and/or circumlocution, and paraphrasing. Strategies involving coherence-detection included identifying text structure, using content schemata, identifying people

and key information, and relying on textual schemata. Finally, monitoring strategies included evaluating comprehension, changing plans, slowing down, rereading, and repeated skimming and scanning. These learner strategies were also classified as comprehension promoting or deterring strategies.

Results revealed that the participants used strategies to almost the same extent when processing texts in both Hebrew and English. Therefore, Sarig (1987) suggested that the learners transferred their L1 reading strategies to the L2. Sarig (1987) also found that the coherence detecting (top-down) strategies led to both successful and unsuccessful reading comprehension in both languages. This finding was in contrast to Block's (1986) results that indicated that general (top-down) strategies led to successful, not unsuccessful, reading comprehension. Sarig's results also indicated that the most language dependent (bottom-up) category of strategies (i.e., the clarification and simplification category) contributed the least of all to overall success in L1 and L2 text comprehension. However, a finer-grained analysis of the differences in strategy use between the higher and lower proficiency students may have led to more complex conclusions as more recent research (Hammadou, 1991; Wolff, 1987) has found differences in strategy use as a function of text difficulty and second/foreign language proficiency level.

In a study of foreign language learners' reading strategies in which verbal report and recall protocols were collected and examined, Wolff (1987) found that strategic processing was affected by learner language proficiency level. He examined the reading strategies of 350 German learners of English varying in age from 12 to 18. After listening to an English narrative read by a native English speaker, the learners were asked to retell

the story in their L1. Then they were asked questions about the cognitive strategies they used to comprehend and recall the story. The recalls and interviews were recorded, transcribed and analyzed for frequency of text related and unrelated propositions, inferences, and strategies reported. On the basis of significant differences in frequencies of propositions and inferences between a subsample of novice (n = 32) and higher proficiency (n = 39) learners, Wolff concluded that the novice English as a foreign language (EFL) learners had a greater tendency to rely on top-down processing than the higher proficiency learners, which he argued was because the novice learners' bottom-up processing was impeded by their language deficiencies; whereas the higher proficiency learners' advanced linguistic knowledge contributed to greater accuracy in their comprehension and use of bottom-up strategies.

Hammadou's (1991) results appear to support Wolff's (1987) conclusions. In Hammadou's study of the relationships between reading comprehension, content knowledge, and the inferencing strategies of beginning and advanced college students studying French (n = 41 beginner; n = 48 advanced) or Italian (n = 43 beginner; n = 34 advanced), she discovered that the beginner foreign language (FL) learners produced significantly longer written recall protocols with more global inferences than the advanced readers. Beginning French readers made 124 inferences, whereas advanced French readers made 111 inferences. Beginning Italian readers made 121 inferences, whereas advanced Italian readers made 70 inferences. This suggests that the beginners used more top-down strategies than the advanced readers.

In a follow-up to her earlier (1986) study, Block (1992) investigated the comprehension-monitoring processes of first and second language readers of English.

Twenty-five college freshmen – 16 proficient readers (8 L1 and 8 L2 readers) and 9 non-proficient readers (3 L1 and 6 L2 readers) – were selected to participate in this study on the basis of their high or low scores on a standardized English reading test. While reading an expository text in English, the participants were asked to identify everything they understood and were thinking as they read each sentence. Results indicated that the L2 readers used similar strategies to those used by the L1 readers. The most proficient readers, regardless of language background, tended to verbalize their strategic plans and check their solutions more often than the less proficient readers did. Not surprisingly, however, the proficient ESL readers did not verbalize as much as the proficient native speakers. This was likely due to native- non-native- speaker differences in the ability to express themselves in English. Results also suggested that when dealing with reading problems, less proficient readers relied more upon local, word-based (bottom-up) processing strategies, whereas the more proficient readers used more global (top-down) meaning-based strategies. These results are in contrast to Wolff (1987) and Hammadou's (1991) results as they found that the less proficient readers used more top-down strategies.

In a later verbal report study, Young and Oxford (1997) explored and compared the strategies employed by 26 female and 23 male native English speaking learners of Spanish while reading two Spanish texts (one edited and one authentic) and one authentic English text. Participants were 14-first year, 14-second year, 9-third year, and 12-fourth year university students of Spanish. Three different kinds of passages were selected from university Spanish textbooks for each level of learners. After reading each passage, participants completed a retrospective think-aloud interview where they reported the

strategies they used to process the passage; retold the passage in an oral recall; and rated their degree of familiarity with the passage topic. Strategies identified in the protocols were coded using a global-local classification scheme. Global strategies included skimming, anticipating content, recognizing text structure, drawing inferences, integrating information, identifying the main idea, speculating beyond the text, and using background knowledge. Local strategies included: stating understanding of words, skipping unknown words, breaking words or phrases into parts, referring to the glossary, using cognates, using syntax, translating a word or a phrase, and paraphrasing.

Results revealed no significant differences between males and females in their (a) use of global versus local strategies, (b) recall scores, (c) passage comprehension ratings, or (d) topic familiarity ratings. However, significant gender differences were found in the use of specific strategies when reading. For example, when reading the Spanish-edited passage, males monitored their reading pace and paraphrased more often than females. Significant differences were also found in the mean number of global versus local strategies reported by the participants when processing both English and Spanish texts. Students relied more upon global strategies when processing the English L1 passage and local strategies when processing the Spanish L2 passage. Overall, these findings support Block's (1992) conclusion that more proficient readers tend to rely on more global, meaning-based (top-down) strategies, whereas less proficient readers tend to rely more upon local, word-based (bottom-up) processing strategies.

In a more recent think aloud study, Liontas (1999) investigated the strategies that 60 native English speaking third-year university learners of Spanish, French, and German used to comprehend and interpret L2 phrasal idioms and authentic texts. Results

suggested that regardless of the L2, readers relied on a combination of both top-down and bottom-up reading strategies. For example, to detect phrasal idioms and comprehend authentic L2 texts readers used word and idiom recognition strategies, lexical access and retrieval strategies, and strategies that relied on contextual and pragmatic support, background and world knowledge, and formal schemata (i.e., knowledge of the organizational structures of different types of texts). These findings support an interactive theory of L2 reading, which maintains that a combination of both top-down and bottom-up strategy use leads to successful L2 reading comprehension.

Unfortunately, the convenience samples and small numbers of participants interviewed in some of the verbal report studies (e.g., Block, 1986, 1992; Sarig, 1987) limit the external validity of the results as the strategies elicited in such studies are likely to be biased in terms of social, educational, and other background factors of the selected samples. For instance, not all individuals may have been able to introspect and verbally report successfully, especially those L2 learners who lacked linguistic knowledge and had to report in the L2. The internal validity of these studies may also have been threatened by researcher influence and the method used to elicit the strategies. For example, the participants may have only focused on certain strategies due to the researcher's probing of these strategies. Persistent probing may have encouraged the participants to make up responses and the act of having to report verbally may have disrupted the participants' normal behaviours. Furthermore, if the participants' thought processes were not readily accessible or easily verbalized this may have inhibited their reporting (Cohen & Scott, 1996). In addition, the use of written recall protocols was also problematic as it is difficult to separate reading comprehension from memory capacity.

Reading Strategy Questionnaire Studies

Self-report questionnaires, in which learners report on their perceptions of their strategy use, are common assessment tools that researchers have also used to investigate second language reading strategies. In a study of foreign language reading, Barnett (1988) investigated the relationships between reading comprehension, strategy use, and perceived strategy use among 272 fourth semester university French students. The students were divided into 19 sections. All sections were taught using a traditional skills-based approach. However, four of these sections received additional training in reading strategies. During the experiment, the students (a) read an unfamiliar text in French, (b) wrote a recall in English, (c) answered background knowledge questions about the topic of the next passage they were to read, (d) completed a rating scale about their strategy use, and (e) answered a 17 item questionnaire designed to elicit the students' perceived strategy use of text- and word-level strategies. Text-level strategies included strategies used to read the passage as a whole (e.g., utilizing background knowledge; considering context; noting the interrelationship of words, actions and ideas; predicting; reading the title; skimming). Word-level strategies included using context to guess word meanings, identifying grammatical categories of words, using reference words, and identifying word families.

The multiple-choice strategy-use questionnaire consisted of what Barnett (1988) identified as effective and less effective text-level and word-level strategies. Effective strategies included attending to passage meaning, attending to grammatical form or function, reading the whole passage once and then again, reflecting on background knowledge of the topic, hypothesizing about what might come next, predicting what the

passage might be about from the title, and guessing what some words mean. Less effective strategies included attending to the meaning of individual words, attending to passage structure, only rereading the difficult sections, not hypothesizing about what comes next, not making connections between paragraphs, and thinking that it is a mistake to skip any words. However, rather than prejudging the effectiveness of these strategies, it would have been valuable if Barnett had asked the students whether they considered the strategies effective or not and why.

Results revealed significant positive relationships between comprehension and strategy use ($r = .38$), comprehension and perceived strategy use ($r = .26$), and strategy use and perceived strategy use ($r = .22$). Students in the high strategy group who used the strategies related to reading through context comprehended better than those in the low strategy group. Students in the high strategy group who perceived that they used the effective strategies scored significantly higher on the comprehension task than the students who perceived that they used the less effective strategies in the low strategy use group. In addition, students who received strategy training showed greater ability to read through context than the students in the control group.

In another self-report study, Carrell (1989) used a questionnaire to (a) investigate second language readers' perceptions of their metacognitive reading strategy use in both their L1 and L2, and (b) examine the relationship between the learners' metacognitive awareness and reading comprehension in both languages (English and Spanish). Two groups of university students participated in the study. Group One comprised 45 native Spanish speakers at intermediate and advanced proficiency levels in English. Group Two consisted of 75 native English university students studying first, second, and third year

Spanish as a foreign language. All participants read two texts in both languages, first in their L2, then in their L1. All four texts dealt with the topic of language. After reading each text, the participants answered 10 multiple-choice comprehension questions followed by a strategy use questionnaire that had the participants rate their level of agreement or disagreement on a scale of 1-5 (from strongly agree to strongly disagree) with 36 statements concerning their reading strategies. The statements about the strategies that the learners used showed how confident they were in their ability to read in each language; what they did when they did not understand something; what they focused on or did to read more effectively; and which reading strategies they found difficult to use in each language.

Carrell (1989) found that (a) the higher proficiency readers perceived global, top-down strategies to be significantly more effective and less difficult to use, and (b) the lower proficiency Spanish FL readers perceived bottom-up, local strategies to be significantly more effective and less difficult to use. Local strategies were "those having to do with sound-letter, word-meaning, sentence syntax, and text details" and global strategies were "those having to do with background knowledge, text gist, and textual organization" (p. 126). Although other researchers (Block, 1992; Hosenfeld, 1977) have also found that successful L2 readers use more top-down than bottom-up strategies, it is difficult to compare the results of these studies because Carrell examined perceived strategy use while Block and Hosenfeld examined strategy use inferred from think aloud data. In addition, all three researchers distinguished proficient from less proficient readers in different ways.

In a more recent study of 128 second year university-level native English speaking learners of German (78 females and 50 males), Schueller (2000, in press) examined the effects of top-down and bottom-up reading strategies instruction on reading comprehension. Three groups of learners (2 treatment, 1 control) were taught using a regular skills-based approach to teaching German. However, while the control group received no special training, one treatment group received training in top-down strategies, and the other treatment group received training in bottom-up strategies. After the strategy training, participants in all three groups read one German text and then completed a written recall in English followed by multiple-choice reading comprehension questions. Results showed that (a) the top-down group outperformed the other two groups on both measures of reading comprehension, (b) the top-down group made more correct inferences in the written recalls than the other two groups, and (c) females outperformed the males on both comprehension measures regardless of strategic training with only one exception: males trained in top-down strategies outperformed the females on the multiple-choice questions.

In another study that utilized a self-report strategy questionnaire, Brantmeier (2000, 2003a, 2003b) examined the relationships between gender, passage content, reading comprehension, and strategy use in a sample of 78 English speaking university students (29 male, 49 female) who were studying intermediate level Spanish. Participants read two Spanish passages, one about boxing and the other about a frustrated housewife. Then they completed a written recall in English, multiple-choice reading comprehension questions, and a strategy use questionnaire consisting of multiple-choice and yes/no questions. Not surprisingly, Brantmeier found a significant interaction between gender

and passage content. Males scored higher on both the written recall and the reading comprehension questions related to the boxing passage, whereas females scored higher on both tasks concerning the frustrated housewife passage. Like Young and Oxford (1997), Brantmeier found no significant gender differences in the overall number of global and local strategies the participants used when reading texts in Spanish as an L2. Results also indicated no significant relationships between global or local strategy use and reading comprehension. This is in contrast to Barnett's (1988) findings where text-level (top-down/global) strategy use was positively related to reading comprehension. However, both males and females used significantly more local than global strategies when reading the Spanish passages.

Although these questionnaire studies provided some evidence of the influence of reading strategies on reading comprehension, the validity of the inferences made from the data collected in each of these studies may have been threatened because the questionnaires or checklists that were used by the learners to report their strategy use may have neglected the strategies that the students actually employed. In this case, the questionnaires could have been incomplete or too simplistic. Another problem associated with these questionnaire studies is that the participants may have made guesses about what they actually did when they read. Consequently, they may have over- or under-estimated the frequency of their strategy use (Cohen & Scott, 1996). In addition, if the questionnaires were not carefully piloted, the questions may have been ambiguous or leading, producing data that may have been difficult to interpret.

Strategy Use and ESL/EFL Test Performance

Language testing researchers are not only interested in the product of reading (e.g., scores on a reading test), they are also interested in studying the processes readers use to solve test tasks. Thus reading strategy research has also been conducted using various combinations of reading strategy questionnaires, verbal reports, written recalls, cloze tests, and standardized multiple-choice reading comprehension tests. Much second language testing research has been concerned with general test-taking or test-wiseness strategies that examinees use when answering multiple-choice and cloze reading items (Allan, 1992; Cohen, 1984; Nevo, 1989; Rogers & Yang, 1996; Yang, 2000). Some examples of the numerous multiple-choice test-wiseness strategies that have been identified in the literature include reading the question first, selecting a choice that is longer/shorter than the others, using the process of elimination to select an answer, and producing an answer to the question before looking at the options (Cohen, 1998). While some second language testing researchers have focused specifically upon test-taking (test-wiseness) strategies, others have examined both reading and test-taking strategies.

In a study of reading comprehension and test-taking strategies, Anderson (1991) examined individual differences in the strategy use of 28 Spanish speaking ESL learners while completing a standardized reading comprehension test and reading two academic texts. Participants were enrolled in a university-level English as a second language program. Their reading comprehension skills were assessed with both forms of the Descriptive Test of Language Skills-Reading Comprehension Test (DTLS) taken at two different times. The test consisted of 15 reading passages followed by two to four multiple-choice comprehension questions for each one, designed to test three types of

reading skills: understanding main ideas, understanding direct statements, and drawing inferences. Students were randomly assigned to two groups. Group One completed Form A of the DTLS and Group Two completed Form B. Approximately one month later, the students (a) read two academic passages taken from freshmen-level texts, (b) answered multiple-choice comprehension questions corresponding to each passage, (c) retrospectively verbalized the strategies they used when reading the passages and answering the questions, and (d) were given 30 minutes to complete the second form of the DTLS while concurrently verbalizing their strategies. The verbal reports were conducted in the participants' language of choice (i.e., Spanish, English, or both languages). The audio-taped protocols were transcribed and coded for strategies in the following five categories: supervising, supporting, paraphrasing, establishing coherence, and test-taking strategies.

Results indicated that (a) the frequency of strategy use in no particular strategy category significantly contributed to success on either the standardized reading comprehension test or the academic reading comprehension questions, and (b) participants who reported using more strategies tended to score higher on the two measures of reading comprehension. However, it is likely that a finer grained statistical item level analysis would have provided greater insights into the relationships between the skills or strategies that the items were designed to measure and the strategies that the learners reported using.

Using the think aloud and standardized reading test data described in the previous study (Anderson, 1991), Anderson, Bachman, Perkins, and Cohen (1991) investigated the relationships among test-taking strategies, item content, and item performance. In

addition to the data collected by Anderson (1991), a content analysis of the items from both forms of the DTLS was conducted. Items were classified in two ways: (a) as implicit or explicit; and (b) as main idea, direct statement, or inference items. Results indicated a significant positive relationship between strategies inferred from the verbal reports and item type. For example, the second most commonly used strategy, trying to match the stem with the text, was reported more frequently for inference type questions than for direct statement questions. Results also revealed a significant relationship between strategy use and item difficulty as fewer strategies were reported for the easy items ($p >$ .67) than for the average ($.33 \leq p \leq .67$) or difficult ($p < .33$) items. The results of this study and the previous study demonstrate the importance of using think aloud data to identify processing strategies that are typically ignored in the interpretation of total test scores. The authors stressed that such procedures should be used to supplement traditional psychometric approaches to construct validation.

In a more recent study of EFL test performance, Storey (1997) examined the processes employed by 25 female Chinese university students when engaged in a 13-item multiple-choice four-option English discourse cloze test. The cloze test was a gap-filling test where words carrying discourse meaning (i.e., words that marked relationships between propositions as opposed to phrase- or clause-bound meaning) were deleted. Participants were selected on the basis of their high oral English proficiency from the English majors enrolled in a two-year teacher education course in Hong Kong. After training in think aloud procedures, the participants verbalized their reading strategies in English while completing the test in a language laboratory. However, it is difficult to determine whether reporting in their L2 limited their analytic reporting abilities.

Although they were all high proficiency learners, they still may have had some difficulties reporting in English.

The think alouds were coded for *macro* (intersentential) and *micro* (intrasentential) discourse processing strategies. The macro-micro strategy classification schema is similar to several of the other binary coding schemas (e.g., global vs. local, top-down vs. bottom-up). Qualitative analyses of the verbal reports revealed that while the deleted discourse markers encouraged the participants to go beyond the sentence level and use *macro* discourse processing strategies (e.g., connecting and inferencing strategies) to process associated arguments and the rhetorical structure of the text, the deleted lexical substitutes merely encouraged *micro* sentence level analysis and surface matching strategies (e.g., paraphrasing or finding words in the passage) despite the test developer's intention to generate gaps that were presumed to elicit integration and connecting strategies. These findings support Storey's argument for the inclusion of concurrent introspective procedures in test validation as think aloud data can provide valuable insights into what the test items actually measure.

In another study of strategy use and test performance, Purpura (1997) examined the influence of EFL learners' cognitive and metacognitive strategy use on their second language test performance. Participants were 1,382 Spanish, Turkish, and Czech high school (60%) and university (40%) high-beginner or above learners of English. They completed an 80 item cognitive and metacognitive strategies questionnaire and the University of Cambridge First Certificate in English (FCE) 70 item Anchor Test. The strategies questionnaire was designed to measure three categories of cognitive processing strategies and two categories of metacognitive strategies. The cognitive processes

included (a) comprehending processes (analyzing inductively, clarifying, inferencing, and translating); (b) memory processes (associating, linking with prior knowledge, repeating, and summarizing); and (c) retrieval processes (applying rules, practicing naturalistically, and transferring). The metacognitive processes included on-line processes (assessing the situation and monitoring), and post-assessment processes (self-evaluating and self-testing).

Of the numerous structural equation models that Purpura (1997) tested, the best statistically and substantively fitting model revealed that metacognitive processes had a significant positive effect on all three cognitive processes; the standardized coefficients were between 0.595 and 0.863. These findings support the theory that metacognitive/control processes oversee and manage the selection, deployment, and evaluation of cognitive processing strategies (Anderson, 1983; Bachman & Palmer, 1996; Gagné et al., 1993). Although metacognitive processing did not directly affect second language test performance, it indirectly affected performance through the process of cognitive retrieval. Purpura (1997) also found that while memory processes had a significant negative effect on grammar ability (standardized coefficient = -0.932), retrieval processes had a significant positive effect on grammar ability (standardized structural coefficient = 0.228). However, the primary limitation in this study was that the self-report questionnaire data collected from the participants' regarding their perceived strategy use may not have directly reflected their actual mental processing behaviours.

In a recent study of EFL reading strategies and test performance, Phakiti (2003) investigated the relationships between cognitive and metacognitive strategy use and reading test performance. A total of 384 Thai university students enrolled in an EFL

course participated in this study. Participants completed an 85 item multiple-choice reading achievement examination followed by a 35 item cognitive and metacognitive strategy use questionnaire. The strategy questionnaire had students rate statements about their learning, test-taking, and reading strategy use on a scale of 1 through 5 (1 = never, 2 = sometimes, 3 = often, 4 = usually, 5 = always). Examples of these statements included "I made short notes or underlined main ideas during the test", and "I used pictures of titles of the texts to help comprehend reading tasks" (p. 55). On the basis of the students' reading test performance and instructor judgments, students were divided into three groups: highly successful, moderately successful, and unsuccessful. Four highly successful and four unsuccessful students were selected for retrospective interviews. These students were asked to retrospectively report in their L1 (Thai) on the strategies they used when completing the EFL reading exam. Then they were asked to complete a 10 minute, six item, multiple-choice reading test, after which they were asked questions about the strategies they used when reading the passage and answering the questions.

Results revealed (a) significant positive relationships between perceived cognitive strategies and reading test performance (r = .39), perceived metacognitive strategy use and reading test performance (r = .47), and cognitive and metacognitive strategies (r = .61); and (b) significant differences in perceived metacognitive and cognitive strategy use among highly successful, moderately successful, and unsuccessful students. The highly successful readers reported using more metacognitive strategies than the moderately successful readers, and the moderately successful readers reported using more cognitive and metacognitive strategies than the unsuccessful readers. Unfortunately, an analysis of the retrospective data revealed that the students mainly focused on describing their test-

taking strategies rather than on their reading strategies. Therefore, trends in actual, rather than perceived cognitive and metacognitive reading strategy use could not be inferred from the data.

Effects of Culture on Reading Strategy Use

In a cross-cultural study of first language reading strategies, Pritchard (1990) found that cultural schemata influenced readers' processing strategies and comprehension levels. Pritchard had 30 American and 30 Palauan proficient 11th grade students think aloud while reading a culturally familiar and a culturally unfamiliar passage in their native languages. After each think aloud, the students were asked to retell the passage. Analysis of the verbal report data revealed 22 processing strategies that Pritchard divided into five categories: developing awareness, accepting ambiguity, establishing intrasentential ties, establishing intersentential ties, and using background knowledge.

When reading the culturally familiar passage, students employed significantly more strategies for establishing intersentential ties and using background knowledge. When reading the culturally unfamiliar passage, students were found to use significantly more strategies for developing awareness and establishing intrasentential ties. Strategies for establishing intersentential ties included relating the stimulus sentence to a previous portion of the text and extrapolating from information presented in the text. Background knowledge strategies included using background knowledge of the discourse format and speculating beyond the information presented in the text. Awareness strategies included recognizing loss of concentration and stating failure to understand a portion of the text, and strategies for establishing intrasentential ties included gathering information and paraphrasing.

Pritchard (1990) also found that (a) the Americans applied a wider variety of strategies (22) than the Palauans (19), (b) the Americans applied strategies with greater flexibility and frequency (3,619) than the Palauans (3,116), and (c) students in both groups recalled significantly more idea units and made fewer distortions of the culturally familiar text than of the unfamiliar text. In addition, it appeared that students who lacked background knowledge of the topic had a tendency to rely more on comprehension monitoring strategies. When processing culturally unfamiliar material, the Americans were more successful at integrating information from the text into their knowledge base than the Palauans, as the Palauans were more likely to focus on word-level strategies. Although this study provided insights into the effects of cross-cultural differences on strategy use, unfortunately Pritchard did not specify which strategies influenced comprehension. More importantly, he made little attempt to explain why the two different groups of students used strategies differently. As Parry (1993) suggested, the knowledge of strategies and when to use them is likely influenced by individuals' experience of text, their written language, and the social process of learning to read.

In another cross-cultural study of reading strategies, Parry (1996) examined the relationship between cultural membership and the EFL reading strategies used by 20 rural Nigerian secondary school students and 25 urban Chinese university graduates when reading academic texts. Parry's research was based upon her own teaching experiences and observations of the roles and processes of L1 and L2 literacy in Yola, Nigeria and Nanjing, China. Different data collection techniques for eliciting information on student reading behaviours were used in these two contexts.

In Nigeria, Parry (1996) administered questions from an outdated school certificate English exam in her English class under exam-like conditions. Then over the next three days, she individually interviewed 20 volunteers about the behaviours they used to comprehend the seven reading passages and to answer the 34 accompanying questions. During the tape-recorded interviews, students (a) read the passages aloud and identified any words that they did not know, (b) answered each question, and (c) explained why they answered as they did. Analysis of student responses revealed that although they correctly answered 28% of the low-level (bottom-up) questions that focused primarily on single lexical items and grammatical knowledge, they correctly answered 41% of the higher-level (top-down) questions that relied upon holistic interpretations of the passages or at least sections of the passages.

In China, Parry (1996) was an instructor of an academic reading and writing course for already certified teachers of English. She had her students "read articles and write essays on four themes: (a) literacy at home, (b) literacy in school, (c) approaching English, and (d) making sense of English text" (p. 676). The teachers' written essays formed the basis for the data analyzed in this part of the study. Seventeen of the 25 teachers chose to write about the strategies they used when reading English texts. Of these 17 teachers, 13 stated explicitly that they concentrated on vocabulary and grammar, which suggested that they placed more emphasis on a bottom-up approach to comprehending English texts than on a top-down approach.

Parry (1996) maintained that the differences between the general tendencies of the two groups reflected differences in their experiences of language and literacy. While the structure of the Chinese EFL textbooks and methods of teaching both Chinese and EFL

tended to encourage bottom-up strategies, the way in which the Nigerians learned to read for meaning and the multilingual environment in which they lived encouraged a global, top-down approach to text. Although the differences in the ways the Nigerian and Chinese students approached English texts may have been due to differences in proficiency, task type, age, and experience, Parry concluded that reading strategies could also "be explained in terms of how different cultural communities represent, use and teach both language and literacy" (p. 687).

Difficulties in Comparing Reading Strategy Studies

Comparisons across all of the reading strategy investigations mentioned thus far are problematic for a number of reasons. First, the participants were quite diverse in terms of their cultural/linguistic backgrounds, educational experiences, second language proficiency levels, and background knowledge. Second, the passages that the participants read varied in genre, topic, length, and difficulty level. Third, a wide variety of reading tasks were used as measures of comprehension (i.e., oral recall, written summaries, multiple-choice questions, cloze tests, and language proficiency and placement tests). Fourth, different criteria were used to distinguish between successful and unsuccessful readers. All of these variables may have differentially affected the participants' reading strategy use and comprehension. Consequently, it is difficult to formulate generalizations about second language reading strategy use. However, in general, the results from these studies suggest that reading comprehension is more likely to occur when individuals use strategies both actively and flexibly during reading.

Second Language Reading Strategy Classification Schemes

As evident from the above discussion, the investigations of second language learners' reading comprehension strategies have produced a wide variety of strategy inventories and classification schemes. Nonetheless, one characteristic that is shared by many of the classification schemes proposed in the L2 literature is that the reading strategies are commonly divided into binary categories. The binary categories are all similar in that they reflect strategies that aid in the comprehension of smaller linguistic units versus those that aid in the comprehension of larger linguistic units. Some of the binary strategy classifications include bottom-up vs. top-down, local vs. global, data driven vs. concept-driven, form-based vs. meaning-based, syntactic vs. semantic, decoding vs. meaning-getting, language-based vs. knowledge-based, word-level vs. text-level, micro vs. macro, analytic vs. synthetic, and analytic and vs. global. Although the terms that are used to refer to either the bottom-up or top-down processes have subtle differences, L2 researchers use these terms together and interchangeably. Thus, it is not uncommon to find statements in the literature such as "novice learners rely primarily on concept-driven (top-down, global) processes when reading texts" (Young & Oxford, 1997, p. 47).

Linguistic Cues

Although over 150 reading comprehension strategies have been identified in the reading literature (Pressley & Afflerbach, 1995), the reading comprehension strategies examined in the current study were limited to a narrow subset of bottom-up and top-down processing strategies that tend to be associated with the specific linguistic cues (discussed

below) that readers attend to when approaching reading comprehension tasks. The five main categories of cues that contribute to a reader's understanding of a text include

1. orthographic/visual cues: individual letters, capitalization, numbers, and pictures or diagrams;
2. morphological/word-level/intrasentential cues: word formation rules and interpretation; the emphasis here is on the smallest units of language that carry information about meaning; making connections within a sentence;
3. syntactic cues: the rules and categories (lexical and functional) that underlie sentence formations;
4. semantic/meaning-based/intersentential cues: how words and sentences are related to the objects (real or imaginary) they refer to and the situations they describe; making connections beyond the sentence; and
5. discourse cues: text organization (titles, subtitles, headings, paragraphs); text types (expository: compare-contrast, cause-effect, problem-solution, description, list; and narrative structures); and genre.

These five categories of linguistic cues are central to theories of reading comprehension and communicative competence as they assist the reader in identifying words and constructing meaning. While low-level cues (orthographic, morphological, and syntactic) tend to aid in the comprehension of smaller linguistic units, high-level cues (semantic and discourse) foster the comprehension of larger units. For example, numbers (orthographic cues) provide information about dates, time, prices, age, or amount of something; the plural –s morpheme (morphological cue) signals more than one; where a word occurs in a sentence (syntactic cue) reflects its lexical (noun, verb, adjective, preposition, adverb) or

functional (auxiliary verb, conjunction, determiner, degree) category (Brown, 1957); the semantic information (semantic cue) carried by the text can constrain the possible interpretations of a passage (Bloom & Wynn, 1997); and the type of text structure (discourse cue) such as an expository compare and contrast structure indicates that the ideas in the text were organized according to similarities and differences (Meyer, 1975).

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

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, and
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

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

These strategies appear in standard classifications employed in one or more of the studies discussed above (see Anderson, 1991; Block, 1986; Carrell, 1989; Phakiti, 2003; Pritchard, 1990; Purpura, 1997; Schueller, 2000, in press; Young and Oxford, 1997).

A Cultural-Historic Approach to Strategy Research

Although culture is often defined as a system of norms and values, ethos and beliefs shared by a particular group of people, according to Storti (1989), it is behaviour that is the principal manifestation and most significant consequence of culture. Learning which reflects the effects of experience on behaviour may therefore be connected to culture as both early life experiences and the common beliefs and values of a person's culture affect the learning process. Research has demonstrated that cultural norms, values, and beliefs are reflected in the educational training that students receive (O'Malley, Chamot, Stewner-Manzanares, Russo, & Kupper, 1985). For example, if rote memorization is highly valued in a culture, then a student's educational training will likely emphasize the development and use of memory strategies. Research also suggests that linguistic background influences the set of strategies examinees bring to a task. For example, research on language learners from multilingual backgrounds indicates that multilingual language learners use a wider variety of strategies and have greater flexibility in strategy use than learners who are predominantly monolingual (Nayak, Hansen, Krueger, & McLaughlin, 1990).

To explain how differences in strategy approach and application are related to both linguistic and cultural differences, researchers have examined how common beliefs and values and their associated behaviours differ from culture to culture. Two main approaches to studying cultural differences have been identified in the literature, a

cultural styles approach and a cultural-historic approach. These two approaches have extremely different foci. While the primary interest of the cultural styles approach is on traits located within groups of individuals that are both independent of tasks and contexts, and are constant over time, the cultural-historical approach centres on the proclivities of people with histories of engagement with specific cultural activities and can thereby account for changes in individuals and practices over time (Gutierrez & Rogoff, 2003).

Rather than speculating about stable traits that are located within individuals who are members of certain ethnic groups, the hypotheses in the current study reflect differences that are related to the collective experiences of people who share the same cultural/linguistic background. Such a perspective avoids an overly simplified explanation as to why individuals or groups act in certain ways and precludes the assumption that certain characteristics are built in to an individual or cultural group and do not change over time. Rather than adopting a cultural styles approach where the Chinese immigrants would likely be characterized as analytic learners and the Arab immigrants as holistic learners on the basis of a learning styles questionnaire, the researcher of the current study adopted a cultural-historical approach where it was assumed that the bottom-up and top-down approaches to processing English used by individuals in these cultural groups were not permanent characteristics of these groups. These strategies were viewed as the dominant strategies that were practised and learned in common cultural/linguistic settings. It was assumed that the tendency to rely on either top-down or bottom-up strategies would lessen as the immigrants became more proficient readers of English and as they were exposed to different methods of teaching English in Canada.

The Influence of Culture, Education, and Language on Strategy Use

Although some cultural and educational factors have been shown to influence strategy preferences (e.g., Bedell & Oxford, 1996; Levine et al., 1996; Harshbarger et al., 1986; Pritchard, 1990; Reid, 1995; Willing, 1988), little explanation has been provided as to why this occurs. In keeping with a cultural-historical approach to characterizing regularities in the variations in reading strategy use among Chinese/Mandarin and Arab/Arabic cultural/linguistic groups, an attempt was made to explain why intermediate Arabic and Mandarin speaking 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 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

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; and
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, 1992).

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 activities focus on developing functional language skills in a learning environment that stresses meaning over form. As Parry (1996) suggests, 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 Arab EFL students receive to communicative activities promotes the development of top-down reading strategies.

Many ESL instructors have noticed that Chinese ESL students tend to use a dictionary more than Arab ESL students. The reason for this differential use is likely reflected in both their linguistic and educational systems. Thompson-Panos and Thomas-Ružić (1983) maintain that Arab students are not highly skilled in using dictionaries when reading and writing because the words in Arabic dictionaries are arranged according to their roots. In English, this would be similar to looking up the word *misconceived* under the entry for *cept* (Thompson-Panos & Thomas-Ružić, 1983). If the educational system does not emphasize the development of such skills when learning the L1, these skills will not be available to transfer to L2 learning, and consequently will not promote a bottom-up approach to reading in an L2. On the contrary, most Chinese students tend to rely heavily upon their dictionaries and as a result usually have well-developed dictionary skills, which encourage the development of a bottom-up approach to reading (Parry, 1996).

The Effects of Linguistic Differences on Strategy Use

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 stated in the introduction, in a cross-linguistic study of ESL reading skills, Fender (2003) 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, 1992). Therefore, one may 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). Since vowels are not usually represented in Arabic orthography, Arabs may be less dependent on local cues in the printed word when reading. If reading in Arabic encourages a reliance on higher-level 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 may rely more upon their background and experiences than upon 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 may cause them to be distracted by less relevant textual information and as a result they may 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, while Chinese word recognition requires extensive orthographic processing skills, alphabetic or consonantal orthographies require a greater connection to phonemes and phonology. Therefore, Chinese ESL learners may 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 (letter-sound) 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 may be attained by students well-prepared on the represented skills or processes that are measured by the reading items but ill-prepared on the under-represented 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 may 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. 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 CLBA item. When comparable examinees do not perform the same on specific test items, the items are said to display DIF. Large DIF indices signify that the items are measuring secondary dimensions that may either be relevant or irrelevant to the construct measured by the test (Roussos & Stout, 1996). 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 which may 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).

DIF on ESL/EFL Proficiency and Placement Tests

Much of the research regarding the effects of language background on second language test performance has been concerned with whether ESL/EFL 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 examinees from different language groups perform differently on such tests at the item level (see Chen & Henning, 1985; Kim, 2001; Ryan & Bachman, 1992; Sasaki, 1991). These studies are discussed below.

Chen and Henning (1985) utilized an adapted Angoff delta-plot method (Angoff & Ford, 1973) to identify DIF items on the UCLA English as a Second Language Placement Exam (ESLPE) across Chinese (n = 77) and Spanish (n = 34) first language groups. The ESLPE consisted of five 30-item subtests: listening, reading, grammar, vocabulary, and writing error correction. Chen and Henning modified the delta-plot DIF detection procedure by plotting difficulty estimates calibrated by the one-parameter IRT model for each item across the two groups on a scatterplot rather than plotting the traditionally used standardized p-values (the proportion of examinees answering the item correctly). The assumption of this modified delta-plot method was that if an item was unexpectedly too difficult for one group and unexpectedly too easy for the other, it would be regarded as exhibiting DIF. Items beyond the 95% confidence interval of the regression line were considered DIF items.

Results indicated that four items favoured the Spanish group. Not surprisingly, the four items were English vocabulary items with close Spanish cognate forms (e.g., the Spanish

cognate for *approximate* is *approximado*). The authors concluded that due to the similarities between English and Spanish, the Spanish speakers had a natural advantage over the Chinese speakers with respect to vocabulary recognition. Since vocabulary was relevant to the construct being measured by the ESLPE, the DIF exhibited by these items may be attributed to an auxiliary dimension of ESL proficiency and deemed benign. However, if the proportion of cognate vocabulary items exceeded the proportion of naturally occurring cognates between the two languages, then the vocabulary subtest would not validly represent the English lexicon. In this case, content representativeness and thus test fairness would become an issue.

Sasaki (1991) conducted a similar study to Chen and Henning's (1985) study in that she also examined DIF in the UCLA ESLPE across Chinese (n = 262) and Spanish (n = 81) language groups. However, she studied a different version of the ESLPE than Chen and Henning, and utilized Scheuneman's chi-square method (Scheuneman, 1979) for detecting DIF in addition to the same modified delta-plot method employed in Chen and Henning's study. Scheuneman's method, like other contingency table approaches, is based on the assumption that after controlling for ability, members of each group are expected to have approximately the same probability of answering each item correctly. To control for ability, Sasaki divided the Chinese and Spanish groups into three ability levels (low, mid, and high) with approximately the same number of students at each level. Then the significance of the differences between observed frequencies and expected frequencies at each of the three ability levels was calculated for each item.

While the modified delta-plot method identified nine DIF items (5 grammar, 4 vocabulary), Scheuneman's method detected 22 DIF items (4 listening, 1 reading, 4 grammar, 7 vocabulary, 6 writing error detection). Substantive analyses of the DIF results indicated DIF

favouring the Spanish group on cognate vocabulary items, and DIF favouring the Chinese group on items containing idiomatic expressions. In both cases, DIF could have been attributed to auxiliary dimensions of ESL proficiency and deemed benign. However, since idiomatic expressions might have been heavily emphasized in the Chinese speakers' instructional backgrounds and not highly emphasized in the Spanish speakers' instructional backgrounds, it was likely that instructional and curricular differences between the two groups had an impact on item performance. Thus, additional investigation into the proportion of idiomatic expression and cognate vocabulary items in the ESLPE is required to address the issue of content representativeness and ultimately test fairness.

Using the Mantel-Haenszel (MH) DIF detection procedure (Mantel & Haenszel, 1959), Ryan and Bachman (1992) examined the extent to which items on the Test of English as a Foreign Language (TOEFL) and the First Certificate in English functioned differentially for equal ability examinees from Indo-European (n = 792) and non-Indo-European (n = 632) L1 backgrounds. Indo-European (IE) examinees were native speakers of French, German, Spanish, and Portuguese, and non-Indo-European (NIE) examinees were native speakers of Japanese, Thai, Chinese, and Arabic. The MH delta difference (MH D-DIF) (Holland & Thayer, 1986) was used to estimate the average amount by which the IE group found a given item more difficult than did comparable members of the NIE group.

On the TOEFL, 32 of the 146 items were found to be easier for the IE group, and 33 items were easier for the NIE group. These differentially functioning items were spread across all three sections of the test (i.e., Listening, Structure and Written Expression, and Vocabulary and Reading). However, on the Listening component, the high (MH D-DIF > 1.5) C level DIF items were not split evenly among the groups as five of the C level DIF items favoured the

NIE group, while only two C level items favoured the IE group. On the First Certificate in English, 25 of the 40 reading and vocabulary items were found to exhibit DIF (13 favoured the IE group, 12 favoured the NIE group). However, eight C level DIF items favoured the IE group while only three C level items favoured the NIE group. The researchers suggested that these differences were not only attributable to differences in the examinees' native languages but also to differences in the examinees' culture and education. Nevertheless, it is not clear whether the DIF may be attributed to auxiliary or nuisance dimensions of ESL proficiency as no substantive analysis of the items was reported in this study.

In a more recent study, Kim (2001) examined DIF across Asian (n = 467) and European (n = 571) language groups on the Speaking Proficiency English Assessment Kit (SPEAK) using the likelihood ratio test (Thissen, Steinberg, & Wainer, 1988) and the ordinal logistic regression approach (Zumbo, 1999). These DIF detection procedures were selected because they were considered appropriate for examining the polytomous scoring scales used in the SPEAK test where grammar, pronunciation, and fluency were rated using an ordinal scale from 0 to 3. Of the three scoring categories examined in this study (i.e., grammar, pronunciation, and fluency), Kim found that both methods yielded similar results in that the grammar and pronunciation scales' discrimination values functioned differentially across the Asian and European groups. While the grammar scale was better at discriminating between the high and low ability European speakers of English, the pronunciation scale was more discriminating for the Asian group. However, the fluency parameter estimates were very similar across the two groups, suggesting that this scale did not show DIF.

In the studies mentioned above, the researchers used a variety of DIF detection methods with diverse populations to examine the extent to which items from ESL placement

and proficiency tests functioned differentially for examinees of equal ability from different first language backgrounds. Although each of these studies provided evidence that linguistic background is one determinant of DIF in ESL test performance, the studies are not without limitations. For instance, the small sample sizes in Chen and Henning's (1985) study and Sasaki's (1991) study may have affected the accuracy of the IRT difficulty parameter estimates. In addition, because the one parameter IRT model assumed constant item discrimination, differences in difficulty among items were confounded with differences in discrimination among items (Camilli & Shepard, 1994). Furthermore, the 95% confidence interval for determining DIF items in both studies was arbitrary. If narrower confidence intervals had been used, more DIF items would likely have been detected. Additionally, the unbalanced sample sizes in Sasaki's study may have inflated Type I error in the Scheuneman chi-square procedure. The primary limitation in Ryan and Bachman's (1992) study was that they did not conduct a substantive analysis of the DIF items identified by the MH procedure. It is likely that a content review of the items may have shed some light on the sources or factors contributing to DIF in the two language groups. Finally, the small number of scoring categories examined in Kim's (2001) study made it difficult to evaluate the comparability of the two DIF detection methods.

Although the statistical methods utilized in these studies were relatively useful for flagging DIF items, to understand the nature of DIF, content analyses were also required to determine why the items functioned differentially between the groups. However, the researchers' attempts to identify the causes of DIF in many of the items using content analyses were not successful. For example, of the 22 DIF items identified by Scheuneman's chi-square method in Sasaki's (1991) study, only four 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), Douglas, Roussos, and Stout (1996) proposed a confirmatory approach to DIF and differential bundle functioning (DBF). This approach, which is based on the Shealy-Stout multidimensional model for DIF (Shealy & Stout, 1993), was used in the current study and is described in the next section.

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, substantive analyses of the test items are conducted in order to generate DIF hypotheses. A DIF hypothesis specifies whether an item or 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/DBF. The second stage in the Roussos-Stout DIF analysis framework involves statistically testing the hypotheses generated in stage one of the analyses. The statistical procedure selected for testing the hypotheses in this study was the Simultaneous Item Bias Test (Stout & Roussos, 1999).

The Simultaneous Item Bias Test

The Simultaneous Item Bias Test (SIBTEST) is a commonly used statistical procedure for detecting DIF. SIBTEST was selected for use in this study for three main reasons. First, SIBTEST has been found to be more effective in detecting DIF than the Mantel-Haenszel and logistic regression procedures (Bolt & Stout, 1996; Ercikan, Gierl, McCreith, Puhan, & Koh, 2002; Gierl, Rogers, & Klinger, 1999; Jiang & Stout, 1998). The identification of an increased number of DIF items may result in a more thorough

analysis of the Canadian Language Benchmarks Assessment (CLBA) Reading Assessment items leading to a more comprehensive evaluation of the test and the reading strategy framework that was used to group the items in this study. Second, SIBTEST uses a regression estimate of the true score, instead of the observed score, to match students on ability, which results in an improved conditioning variable. Third, SIBTEST can be used to test bundles of DIF items. DBF analyses increase statistical power and reduce the number of statistical tests, thereby controlling Type I error (Nandakumar, 1993).

Shealy and Stout (1993) provide a comprehensive and technical discussion of the SIBTEST procedure. SIBTEST can be used to test DIF hypotheses and quantify the size of DIF by estimating a measure of the effect size ($\hat{\beta}_{\text{UNI}}$) for each item and bundle (Stout & Roussos, 1995). In the SIBTEST procedure, items on the test are divided into two subsets, the studied or suspect subtest and the matching subtest. The studied subtest contains the item or bundle of items believed to measure the primary and secondary dimensions identified in the substantive analysis, whereas the matching subtest contains the items believed to measure only the primary dimension. In other words, the studied subtest contains items that are suspected of having DIF, while the matching subtest contains items that ideally have no DIF. The matching subtest places the reference group, which is the favoured group of examinees, and the focal group, which is the disadvantaged group of examinees, into subgroups at each score level so their performance on items from the studied subtest can be compared.

The amount of DIF in the studied subtest is reflected in the effect size estimate $\hat{\beta}_{\text{UNI}}$, which is 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 0 and standard deviation of 1 under the null hypothesis of no DIF. The statistical hypothesis tested by SIBTEST is

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

versus

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

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

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

where $\hat{\sigma}(\hat{\beta}_{UNI})$ 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 $|SIB| > z_1 - \frac{\alpha}{2}$. A statistically significant value of $\hat{\beta}_{\text{UNI}}$ that is positive indicates DIF/DBF

against the focal group and a negative value indicates DIF/DBF against the reference group.

Roussos and Stout (1996) adopted the ETS guidelines for classifying DIF as negligible, moderate, or large. Therefore, the following $\hat{\beta}_{UNI}$ values obtained from a single-item SIBTEST analysis are used for classifying DIF:

- No DIF: Null hypothesis is not rejected and $|\hat{\beta}_{UNI}| \cong 0$,
- Negligible or Level A DIF: Null hypothesis is rejected and $|\hat{\beta}_{UNI}| < 0.059$,
- Moderate or Level B DIF: Null hypothesis is rejected and $0.059 \leq |\hat{\beta}_{UNI}| < 0.088$,
- Large or Level C DIF: Null hypothesis is rejected and $|\hat{\beta}_{UNI}| \geq 0.088$.

Shealy and Stout (1993) provide a comprehensive and technical discussion of the SIBTEST procedure. Unfortunately, however, no guidelines for classifying DBF are currently available.

Literature Summary

Six main conclusions may be drawn from the literature discussed in this chapter. First, although the L2 reading strategy investigations have produced a wide variety of strategy classification schemes, one characteristic that is shared by several of the strategy inventories proposed in the L2 literature is that the reading strategies are commonly divided into binary categories, which reflect local, bottom-up and global, top-down processing strategies. Thus, this binary reading strategy framework was used as the basis for classifying the reading strategies in the current study. Second, despite conflicting results regarding which of these two categories contribute the most to reading comprehension, the common conclusion from studies of the relationship between strategy

use and reading ability is that reading comprehension is more likely to occur when individuals use strategies both actively and flexibly during reading (i.e., they use appropriate strategies from either category given the nature of the context). Therefore, a clear understanding of reading strategies is necessary to help learners discover when, where, and how to use strategies effectively. Third, the results from the strategy training studies suggest that reading strategy training improves comprehension. Thus, before successful reading strategy training programs specifically designed for language learners from different linguistic/cultural groups can be developed, researchers need to explore the differences in strategy use that exist between different linguistic/cultural groups. Fourth, the transfer study results indicate that when reading academic texts, readers appear to transfer their L1 strategies to the L2 as they use similar strategies when reading in both languages. Consequently, one may assume that Arabic and Mandarin speaking ESL learners will transfer their L1 strategies to the context of learning ESL. Fifth, results from the cross-cultural literature show that cultural familiarity with the text topic affects strategy use and EFL learners from different cultures and educational backgrounds tend to rely on different reading strategies and different word recognition strategies when attempting English academic reading tasks. As a result, one may expect that Arabic and Mandarin speaking examinees will rely on different reading strategies when approaching reading comprehension tasks. Finally, due to the lack of success researchers have had when trying to interpret single-item DIF statistics, it is necessary to link both substantive and statistical methods in a DIF analysis framework as this will promote a better understanding of the nature of DIF/DBF on ESL placement and proficiency tests. Hence, a confirmatory approach to DIF/DBF was used in the current study to compare

examinees' scores from Arab and Chinese cultural/linguistic groups on the CLBA reading items.

As mentioned in the introduction, three research questions were investigated in this study. In Part 1 of the study, question 1 was addressed by the verbal report data collected from intermediate Arabic and Mandarin speaking ESL learners as they worked through Form 1, Stage II of the CLBA Reading Assessment. In Part 2 of the study, questions 2 and 3 were addressed by applying the Roussos-Stout (1996) multidimensionality based DIF analysis framework to the study of Arabic and Mandarin speaker differences in ESL reading strategies on the CLBA reading items.

Research Questions:

1. What are the bottom-up and top-down reading strategies that intermediate proficiency Arabic and Mandarin speaking ESL learners employ when reading and answering the CLBA reading items?
2. Is there evidence for differential item performance for Arabic and Mandarin speaking examinees on the CLBA Reading Assessment?
3. If so, is the source of differential performance related to differences in reading strategy use?

More specifically, it was hypothesized that the Arabic speakers would outperform the Mandarin speakers on the items and bundles of CLBA reading items that were presumed to elicit top-down reading strategies, while the Mandarin speakers would outperform the Arabic speakers on the items and bundles of items that were presumed to elicit bottom-up reading strategies. The next chapter describes the methodology that was used to address these questions and hypotheses.

CHAPTER III: METHOD

Overview

The current study was undertaken to develop a theoretical framework to test the hypothesis that the bottom-up items included in the Canadian Language Benchmarks Assessment (CLBA) reading subtest favour Mandarin speakers while the top-down items favour Arabic speakers. The study was conducted in two parts. In the first part, verbal report data were collected from Arabic and Mandarin speaking intermediate ESL learners as they completed the CLBA Reading Assessment. The verbal reports were conducted to clarify and elaborate on the bottom-up and top-down reading strategy classification schema that the expert raters then used to classify the CLBA items in the second part of this study.

In Part 2 of the study, a comprehensive analysis of the CLBA reading items was conducted using substantive and statistical methods. Two samples of examinees were drawn from previously administered CLBA Form 1, Stage II Reading Assessments. One sample consisted of 250 Arabic speaking immigrants, and the other consisted of 250 Mandarin speaking immigrants. Three ESL reading experts classified each of the 32 CLBA reading items into one of the seven bottom-up or five top-down reading strategy categories that had emerged from the data in Part 1 of this study. The item reviewers were also asked to rate how important each strategy was in answering each question. Differential item and bundle functioning analyses were then conducted to determine whether items and groups of CLBA items (classified according to the bottom-up, top-down organizing principle), functioned differentially for equal ability Arabic and Mandarin ESL learners.

Method

Part 1 - Verbal Report

Participants

Arabic and Mandarin speaking immigrants were recruited from intermediate ESL college classes to participate in Part 1 of this study. A letter written to invite students to participate in the study was given to the instructors to hand out to Arabic and Mandarin speaking students in their intermediate ESL classes (see Appendix A). The names of the students who expressed an interest in the study were forwarded by the teachers to an administrative assistant who coordinated the student recruitment and scheduling of participants at the college. Only those students who were literate in their L1 (i.e., had at least 11 years of basic education in their country of origin); who had reached a language threshold in English (i.e., had mastered the basic vocabulary and decoding skills required for placement in an intermediate ESL class); and who had not resided in Canada for more than two years were selected.

Sample size was determined by data saturation. Saturation occurs when no new or useful information about the categories can be obtained (Glaser, 1978; 1992; Glaser & Strauss, 1967). Researchers suggest that data saturation is typically reached after the analysis of 5 to 10 protocols (Conrad, 1978; Glaser & Strauss, 1967; Jones, 1980; Rennie, 1984). To clarify and elaborate on the reading strategy 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.

Although it appeared that saturation was complete after 5 Mandarin participants' and 4 Arabic participants' verbal reports had been collected, transcribed, coded and re-coded, to ensure data saturation had been achieved, verbal reports were collected from 3 additional Arabic participants and 3 additional Mandarin participants. While 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 coding schema.

Instruments

Background Questionnaire. All participants were interviewed using a personal background questionnaire to obtain information on their first language, age, gender, education level, country of birth, length of time studying English, and languages spoken and studied (see Appendix B). To ensure that the participants understood the questions, bilingual translators assisted with the interviews. Participants' responses were recorded on the questionnaire at the time of administration.

Canadian Language Benchmarks Assessment - Reading Assessment. The participants were assessed with the CLBA Reading Assessment, which is a reading comprehension test that requires examinees to attempt a range of different task types. In the past, CLBA Assessment results have predominantly been used to place immigrants in appropriate English as a second language classes. However, CLBA results are currently also being used as a means of establishing admissible levels of English language proficiency in some post-secondary institutions.

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, a content analysis conducted by the researcher revealed that Stage I of the CLBA Reading Assessment primarily comprises bottom-up questions that mainly test vocabulary knowledge and transcoding rather than reading comprehension, whereas Stage II elicited 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, 2000, in press; Young & Oxford, 1997) were used as the starting point in this preliminary classification of the items. The second reason for selecting Form 1, Stage II was that the minimum sample size requirement of 250 Arabic speakers who had completed Stage II was satisfied with Form 1.

Form 1, Stage II of the CLBA Reading Assessment consists of eight dichotomously scored constructed-response items and 24 multiple-choice, four-option items. The items follow four passages (Tasks A-D), which represent four different genres and range in length from 251 to 547 words.

Reading Ability, Preferences, and Strategies Questionnaire. This supplementary questionnaire was designed to obtain self-report data on the participants' perceived reading abilities, reading preferences, and perceived strategy use while answering the CLBA reading questions and when reading silently in English (see Appendix C). Carrell's (1989) L2 reading questionnaire was used as a guideline for constructing the perceived strategy use questions. The participants used a 5-point Likert scale (1 = strongly disagree, 2 = disagree, 3 = unsure, 4 = agree, and 5 = strongly agree) to rate their responses. This questionnaire was administered after the verbal reports had been completed. To ensure that the participants understood the questions, bilingual translators assisted with the

administration of the questionnaire. Participants' responses were recorded on the questionnaire at the time of administration.

Procedure

Ericsson and Simon (1993) developed a model for verbalization processes of subjects under certain conditions so one can make inferences about the cognitive processes that produce the verbalization. They distinguished between two types of verbalization: concurrent and retrospective. The type of directions and the time lapse between the task performance and the verbalization determines the method of verbalization. Concurrent verbalization is obtained as the information is attended to during task performance. Retrospective verbalization occurs after the task has been completed. Verbal protocol analysis (VPA) is different from other verbal analysis such as discourse or content analysis in that the other analyses focus on what is said, whereas VPA makes inferences about the cognitive processes that produce the verbalizations. The protocol comprises the utterances made by the individual and the set of protocols constitute a body of qualitative data.

VPA was used in this study to elaborate on and clarify the coding categories that were used to test the hypotheses about the cognitive processing strategies that differentiate Arabic speaking ESL learners from Mandarin speaking ESL learners in Part 2 of the study. The verbal report procedures applied in this study follow the initial model suggested by Ericcson and Simon (1993), which was further refined by Pressley and Afflerbach (1995), and then again by Green (1998). The VPA steps outlined in Green (1998) were followed in the current study. These steps are described below.

Procedure selection. Concurrent and retrospective non-mediated reports were chosen to avoid the possibility that researcher probes could lead the participants. Thus, participants were asked to report both concurrently and retrospectively as they worked through the CLBA reading items.

Training of bilingual interpreters. Prior to testing the verbal report participants, the bilingual interpreters (a bilingual speaker of Arabic and English, and a bilingual speaker of Mandarin and English) were trained in the verbal report procedures and asked to read and sign a confidentiality agreement (see Appendix D).

Verbal report procedures and training of participants. Each participant met with the researcher and a bilingual interpreter in an empty office at the college. The participant sat at a table on which there were two microphones and a folder containing the experimental materials. These materials consisted of a consent form (see Appendix E), a background questionnaire, a sheet of directions, a practice passage and questions (i.e., CLBA Reading subtest Form 2, Stage II, Task A), and the CLBA Reading subtest Form 1, Stage II, Tasks A-D.

To reduce the cognitive load on the participants, verbal report data were collected on two different afternoons during the same week (i.e., Tuesday and Thursday, or Wednesday and Friday). On day one, after the participant had read and signed the consent form, the researcher and a bilingual interpreter interviewed the participant using the background questionnaire. Then verbal reports were conducted in the student's language of choice (i.e., his/her L1, English, or both languages) to identify the strategies he/she used when answering the CLBA reading items.

Initially, each participant was introduced to concurrent and retrospective verbal report procedures and provided with a chance to practice his/her verbal reporting skills with four or more reading comprehension questions taken from Form 2, Stage II of the CLBA. Each participant practiced reporting 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 seconds, they were reminded to keep talking. During this training session, it was emphasized that they verbalize whatever was going through their minds in whatever form it occurred as they attempted to complete the reading tasks. This produced the concurrent think aloud data. Then after one of four possible multiple-choice options was selected, the participants were asked to report what they could remember about what they were thinking and what words they were attending to from the time they read each question until they selected an answer. This produced the retrospective data.

The instructions for familiarizing the participants with the verbal reporting techniques and practicing the procedures are specified in Appendix F. The validity of the technique was maximized by ensuring that the procedures were adhered to (i.e., standardized instructions were used; the instructions discouraged the participants from rationalizing their thoughts; there was minimum researcher/interpreter intervention; and both concurrent and retrospective verbal reports were collected to maximize the amount of information obtained regarding the participants' reading strategy use).

Once the participants were accustomed to the verbal report procedures, they were administered the first 14 CLBA reading comprehension questions from Form 1, Stage II, and prompted to think aloud while completing each question, and then report retrospectively after completing each question. The participants' responses were audio-

taped and subsequently translated into English where necessary, and transcribed for analysis. To distinguish the parts where the participants responded in English from the parts where they responded in their L1, the responses made in the L1 were typed in square brackets. Short pauses of less than 5 seconds were indicated in the transcripts by three dots. Longer pauses were indicated by five dots and were followed by the researcher's/interpreter's instruction to keep talking.

Supplementary data collection. Upon completion of the questions corresponding to each passage, the participants were asked to (a) rate their familiarity of the topic using the scale (1 = not familiar, 2 = familiar, 3 = very familiar); and (b) rate their comprehension of each of the items and the passage using the 5-point Likert-type scale (1 = none of it, 2 = some of it, 3 = half of it, 4 = most of it, 5 = all of it). The purpose of the topic familiarity ratings was to determine whether there were differences in the background knowledge that the Arab and Chinese participants brought to the text.

Day two data collection procedures. On day two, using the verbal report procedures specified above, participants were asked to complete the remaining 18 CLBA Form 1, Stage II items. After completing the verbal reports, participants were interviewed using the strategies questionnaire.

Data Analysis

Segmenting and coding the protocols. The protocols were segmented and coded by the researcher for types of bottom-up and top-down reading strategies. Each segment of the protocols 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.

After each new participant's data were collected, transcribed, segmented, and coded, the reading strategy coding schema was revised and the previously coded protocols were recoded using the modified strategy classification scheme. By categorizing and re-categorizing the strategies that emerged from the readers' protocols, structure was imposed upon the data, creating a focused search for bottom-up and top-down reading strategies.

Interrater agreement. Consistency of the coding was investigated by having an independent rater code 11 of the 32 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 3 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.

Frequencies of bottom-up and top-down reading strategies. Coded data were quantified and the frequency of Arabic and Mandarin speakers' reading strategies was examined. The strategies identified in the verbal reports formed the coding schema that was used in Part 2 of this study. This ensured that the strategic processes the test-takers

engaged in were reflected in the coding schema that the reading experts used to code the CLBA reading items in Part 2 of the study.

Supplementary data analysis. The data from the background and strategies questionnaires, and the familiarity and comprehension ratings were entered into SPSS 12.0, verified for 100% accuracy, and analyzed to determine the participants' demographic characteristics, and their CLBA reading topic familiarity, perceived passage and item comprehension, and their strategy preferences.

Method

Part 2 - Differential Item and Bundle Functioning

Sample

After the researcher and an assistant had signed confidentiality agreements with the two immigrant referral centres in the province of Alberta where the CLBA performance data were collected (see Appendixes G and H), item level data from 250 previously administered CLBA Form 1, Stage II Reading Assessments in each of the first language groups (i.e., 250 Mandarin and 250 Arabic speakers' assessments) were entered into SPSS 12.0 and verified for 100% accuracy. In an attempt to control for first and second language proficiency, only those ESL learners who had (a) completed at least 11 years of education in their L1, and (b) completed both stages of the CLBA Reading Assessment were included in this sample. It was assumed that learners with this minimum education level would have well developed reading skills and strategies in their L1 and would not have difficulties reading in their L1. It was also assumed that learners who had completed Stage II would have mastered basic decoding skills and basic vocabulary in English. The need to control for L1 and L2 linguistic proficiency is reflected in Alderson's (1984) view that the

skills, strategies, and knowledge from the L1 can only be transferred to L2 reading if the reader has attained a certain level of proficiency in the L2. Also, to control for the influence of ESL education, only the initial assessments of immigrants who had spent up to two years or less than two years in Canada at the time of testing were included in the sample.

Demographic data on the Arabic and Mandarin speakers' gender, age, education, and length of residence in Canada at the time of testing were also collected, entered into SPSS 12.0, and verified for 100% accuracy.

Procedure

Following the recommendations made by Douglas, Roussos, and Stout (1996), a confirmatory approach to DIF was used to examine and statistically test items and groups/bundles of items that were presumed to elicit performance differences for equal ability native Arabic and Mandarin speakers on the CLBA reading items. The theoretical framework, hereafter referred to as the reading classification schema or reading strategy framework, developed in Part 1 of this study, provided the organizing principle for grouping the test items together in terms of bottom-up and top-down strategies so that the effects of first language and cultural background on differential item and bundle functioning could be examined. The assumption was made that if the question was answered correctly, the reader would have focused on the appropriate passage and item cues and successfully employed the predicted strategy that the experts believed was elicited by the item.

Three ESL experts used the reading classification schema to code the 32 reading items included in Form 1, Stage II, of the CLBA. After signing a consent form (see

Appendix I) and participating in a training session which introduced the raters to the coding schema, each rater was asked to independently classify the questions into the bottom-up and top-down reading strategy categories. Thus the items were coded according to the reading strategy that the expert judges 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 they could reach a consensus on the item codings which they disagreed. Then the items were grouped into bundles based on the consensus codes and the following hypotheses were tested using the CLBA item level data:

H_1: The Arabic ESL learners will outperform the Mandarin ESL learners on the bundles of items that rely on top-down processing strategies.

H_2: The Mandarin ESL learners will outperform the Arabic ESL learners on the bundles of items that rely on bottom-up processing strategies.

Although the statistical procedure used in Part 2 of the study (i.e., SIBTEST) is based on the assumption that each item can be classified according to one dominant category, reading comprehension items often contain several different characteristics that may elicit diverse strategies. Therefore, in addition to identifying the most important or salient strategy for each CLBA item, the expert raters were also asked to rate how important each of the 12 reading strategies was in answering each question using the following scale: 1 = not at all salient; 2 = not very salient; 3 = salient; and 4 = very salient (see the rater coding sheet in Appendix J).

Data Analysis

The computer program titled Simultaneous Item Bias Test (SIBTEST; Stout & Roussos, 1999) was used to determine which items and bundles of items displayed statistically significant differential item functioning (DIF), and differential bundle functioning (DBF). A four-step analysis (see Gierl, Bisanz, Bisanz, Boughton, & Khaliq, 2001) was used to test the reading strategy hypotheses. First, a single-item SIBTEST analysis was conducted to provide effect size measures ($\hat{\beta}_{\text{UNI}}$ values) for each of the items. In this analysis, each item was treated as a single-item studied subtest, while the remaining items served as the matching subtest. Second, the reading classification schema was used as the organizing principle to group and graph the $\hat{\beta}_{\text{UNI}}$ values for the items into the bottom-up and top-down strategy categories. Third, the graph was visually inspected to identify patterns in the way the items in the strategy categories were functioning. Fourth, the bundles of items in each strategy category were tested at an alpha level of 0.05. To ensure that the matching subtest was a homogeneous measure across the two groups, the matching subtest for the top-down bundles consisted of the 18 bottom-up test items, and the matching subtest for the bottom-up bundles consisted of the 14 top-down test items. These analyses showed whether there were systematic ways in which the two linguistic/cultural groups responded to groups of test items that were presumed to measure the common secondary dimensions of bottom-up and top-down reading strategies.

CHAPTER IV: RESULTS

Part 1 - Verbal Report

Background Questionnaire

Results from the Background Questionnaire (Appendix B) indicated that the 4 female and 3 male Arabic speaking participants immigrated to Canada from seven different countries: Egypt, Jordan, Kuwait, Lebanon, Libya, Sudan, or Syria. Demographic information is provided in Table 1. The mean length of time spent studying English was 5 years and 4 months in their home countries, and 9 months in Canada. At the time of testing, the Arabic speaking participants had lived in Canada between 7 and 24 months (M = 16.29 months). The mean age of the Arabic speakers was 32.86 (Mdn = 32.0). All of these participants had between 12 and 19 years of education (M = 14.86). In addition to being fluent in oral and written Arabic, 2 of these participants reported that they were also fluent in a second language, either French or Baria. The female participants' occupations included student, secretary, kindergarten teacher, and zoologist. The male participants' occupations included civil engineer, electrical engineer, and electrician.

The Mandarin speaking participants consisted of 4 males and 4 females who had immigrated to Canada from mainland China. The mean length of time spent studying English was 8 years and 5 months in China, and 10 months in Canada. At the time of testing, these participants had lived in Canada between 6 and 24 months (M = 13.43). The mean age of the Chinese participants was 36.13 (Mdn = 36.5). Their years of education ranged from 15 to 23 years (M = 16.94). None of the Chinese participants was fluent in a second language. The female participants' occupations included instrument designer,

statistician, oilfield engineering assistant, and customer service representative. The male participants' occupations included mechanical engineer, software engineer, electronic engineer, and surgeon.

Table 1

Demographic Information for the Verbal Report Participants in Part 1 of the Study

Demographic Variables	Arabic $n = 7$			Mandarin $n = 8$		
	M	*SD*	*Range*	*M*	*SD*	*Range*
Time spent studying English in their home country (years)	5.32	5.87	0-15	8.39	2.39	5-11
Time spent studying English in Canada (months)	9.00	7.66	3-24	10.00	6.99	4-24
Length of residence in Canada (months)	16.29	7.54	7-24	13.43	7.57	6-24
Age (years)	32.86	7.71	21-45	36.13	7.40	25-48
Education (years)	14.86	2.41	12-19	16.94	2.76	15-23

Canadian Language Benchmarks Assessment - Reading Assessment

In Part 1 of the study, the mean CLBA Reading Assessment score for the Arabic speakers was 13.43 out of 32 (*Mdn* = 15, *SD* = 3.87, Range = 8-19), whereas the mean score for the Mandarin speakers was 21.63 (*Mdn* = 22.5, *SD* = 3.89, Range = 16-26).

CLBA Reading Assessment Topic Familiarity, and Perceived Passage and Item Comprehension

After completing the CLBA questions for each CLBA task, the participants in Part 1 of the study were asked to (a) rate their familiarity of the topic using the scale (1 =

not familiar, 2 = familiar, 3 = very familiar); and (b) rate their comprehension of each of the items and the passage using the 5-point Likert-type scale (1 = none of it, 2 = some of it, 3 = half of it, 4 = most of it, 5 = all of it). These results are presented in the following subsections titled topic familiarity, passage comprehension, and item comprehension.

Topic familiarity. The distributions of the Arabic and Mandarin speakers' responses regarding their familiarity with the four CLBA reading passage topics are presented in Table 2. A greater proportion of Arabic speakers than Mandarin speakers

Table 2

Distributions of CLBA Topic Familiarity by Language Group in Percent (n): Arabic Speakers (n = 7), Mandarin Speakers (n = 8)

CLBA Reading Task	Arabic Not Familiar	Arabic Familiar	Arabic Very Familiar	Mandarin Not Familiar	Mandarin Familiar	Mandarin Very Familiar
Task A	0	85.7 (6)	14.3 (1)	37.5 (3)	62.5 (5)	0
Task B	57.1 (4)	28.6 (2)	14.3 (1)	25.0 (2)	62.5 (5)	12.5 (1)
Task C	100 (7)	0	0	100 (8)	0	0
Task D	28.6 (2)	57.1 (4)	14.3 (1)	50.0 (4)	50.0 (4)	0

expressed that they were familiar or very familiar with the topics of Tasks A and D: 100% of the Arabic speakers as opposed to 62.5% of the Mandarin speakers responded that they were either familiar or very familiar with the topic of Task A; and 71.4% of the Arabic as opposed to 50.0% of the Mandarin speakers responded that they were either familiar or very familiar with the topic of Task D. In contrast, 75.0% of the Mandarin speakers as opposed to 42.9% of the Arabic speakers reported that they were either

familiar or very familiar with the topic of Task B. However, no participants in either of the groups were familiar with the topic of Task C.

Passage comprehension. The distributions of the Arabic and Mandarin speakers' responses regarding their understanding of the four CLBA reading passages are presented in Table 3. For Tasks A, B, and C, the Mandarin speakers indicated that they understood a greater proportion of the passages than the Arabic speakers. However, while all of the Arabic speakers and 7 of the Mandarin speakers reported that they understood 50.0% or more of the Task D passage, 1 Mandarin speaker indicated that he comprehended less than 50.0% of the passage.

Table 3

Distributions of CLBA Passage Comprehension by Language Group in Percent (n): Arabic Speakers (n = 7), Mandarin Speakers (n = 8)

CLBA Reading Task	Arabic Less than half	Arabic Half	Arabic More than half	Mandarin Less than half	Mandarin Half	Mandarin More than half
Task A	0	14.3 (1)	85.7 (6)	12.5 (1)	0	87.5 (7)
Task B	14.3 (1)	14.3 (1)	71.4 (5)	0	0	100 (8)
Task C	57.1 (4)	42.9 (3)	0	12.5 (1)	12.5 (1)	75.0 (6)
Task D	0	57.1 (4)	42.9 (3)	12.5 (1)	25.0 (2)	62.5 (5)

Note. Participants were asked to indicate their comprehension of the passage on a scale of "none of it" (1) to "all of it" (5); scores of 1 and 2 were combined into "less than half," while scores of 4 and 5 were combined into the "more than half" category.

Item comprehension. The distributions of the Arabic and Mandarin speakers' responses regarding their understanding of the 32 CLBA reading comprehension items are presented in Table 4. All of the Arabic speakers reported that they understood most or all of the words in 12 of the items, while all of the Mandarin speakers reported that they understood most or all of words in 11 of the items. One Arabic speaker reported that she understood less than half of the vocabulary in three of the items (Task C, items 2 and 4; and Task D, item 2); and 2 Arabic speakers reported that they understood less than half of the vocabulary in one of the items (Task D, item 8). One Mandarin speaker reported that he understood fewer than half of the words in four of the questions (Task A, item 4; Task B, item 2; and Task D, items 3 and 8); and 2 Mandarin speakers reported that they understood less than half of the vocabulary in one of the items (Task D, item 2). Two of the items that contained difficult vocabulary for the Arabic speakers also presented difficulty for the Mandarin speakers (Task D, items 2 and 8). Therefore, although the two groups generally had difficulty with the vocabulary in different items, both groups had difficulty with the vocabulary in a similar number of items.

Reading Ability, Preferences, and Strategies Questionnaire

Perceived reading ability. Table 5 shows the participants' perceptions of their reading abilities in English and their L1. Although 6 of the participants in both language groups were either unsure of their English reading abilities or they disagreed with the statement "you are a good reader in English," all of the participants reported that they were good readers in their first language.

Reading preferences. Table 5 also shows the participants' preferences for learning English. Five (71.4%) of the Arabic speakers and 4 (50.0%) of the Mandarin speakers

Table 4

Distributions of CLBA Item Comprehension by Language Group in Percent (n): Arabic Speakers (n = 7), Mandarin Speakers (n = 8)

CLBA Reading Task	Arabic Less than half	Arabic Half	Arabic More than half	Mandarin Less than half	Mandarin Half	Mandarin More than half
Task A Item 1	0	42.9 (3)	57.1 (4)	0	12.5 (1)	87.5 (7)
Task A Item 2	0	0	100 (7)	0	12.5 (1)	87.5 (7)
Task A Item 3	0	14.3 (1)	85.7 (6)	0	12.5 (1)	87.5 (7)
Task A Item 4	0	0	100(7)	12.5 (1)	12.5 (1)	75.0 (6)
Task A Item 5	0	0	100 (7)	0	12.5 (1)	87.5 (7)
Task A Item 6	0	0	100 (7)	0	25.0 (2)	75.0 (6)
Task B Item 1	0	0	100 (7)	0	0	100 (8)
Task B Item 2	0	0	100 (7)	12.5 (1)	12.5 (1)	75.0 (6)
Task B Item 3	0	0	100 (7)	0	0	100 (8)
Task B Item 4	0	0	100 (7)	0	12.5 (1)	87.5 (7)
Task B Item 5	0	28.6 (2)	71.4 (5)	0	25.0 (2)	75.0 (6)
Task B Item 6	0	0	100 (7)	0	0	100 (8)
Task B Item 7	0	14.3 (1)	85.7 (6)	0	0	100 (8)
Task B Item 8	0	0	100 (7)	0	12.5 (1)	87.5 (7)
Task C Item 1	0	42.9 (3)	57.1 (4)	0	0	100 (8)
Task C Item 2	14.3 (1)	28.6 (2)	57.1 (4)	0	12.5 (1)	87.5 (7)
Task C Item 3	0	42.9 (3)	57.1 (4)	0	12.5 (1)	87.5 (7)
Task C Item 4	14.3 (1)	14.3 (1)	71.4 (5)	0	12.5 (1)	87.5 (7)
Task C Item 5	0	14.3 (1)	85.7 (6)	0	12.5 (1)	87.5 (7)
Task C Item 6	0	14.3 (1)	85.7 (6)	0	12.5 (1)	87.5 (7)
Task C Item 7	0	0	100 (7)	0	12.5 (1)	87.5 (7)
Task C Item 8	0	14.3 (1)	85.7 (6)	0	0	100 (8)
Task D Item 1	0	14.3 (1)	85.7 (6)	0	0	100 (8)
Task D Item 2	14.3 (1)	14.3 (1)	71.4 (5)	25.0 (2)	12.5 (1)	62.5 (5)
Task D Item 3	0	14.3 (1)	85.7 (6)	12.5 (1)	0	87.5 (7)
Task D Item 4	0	14.3 (1)	85.7 (6)	0	0	100 (8)
Task D Item 5	0	14.3 (1)	85.7 (6)	0	0	100 (8)
Task D Item 6	0	14.3 (1)	85.7 (6)	0	0	100 (8)
Task D Item 7	0	0	100 (7)	0	0	100 (8)
Task D Item 8	28.6 (2)	14.3 (1)	57.1 (4)	12.5 (1)	0	87.5 (7)
Task D Item 9	0	14.3 (1)	85.7 (6)	0	25.0 (2)	75.0 (6)
Task D Item 10	0	28.6 (2)	71.4 (5)	0	25.0 (2)	75.0 (6)

Note. Participants were asked to indicate their understanding of the passage on a scale of "none of it" (1) to "all of it" (5). Due to the lack of extreme scores, scores of 1 and 2 were combined into "less than half," whereas scores of 4 and 5 were combined into the "more than half" category.

Table 5

Percentage of Perceived L1 and L2 Reading Ability and Preferences for Learning English by Language Group (n): Arabic Speakers (n = 7), Mandarin Speakers (n = 8)

Ability

You are a good reader in English.

	Disagree	Unsure	Agree
Arabic	0	85.7 (6)	14.3 (1)
Mandarin	25.0 (2)	50.0 (4)	25.0 (2)

You are a good reader in your first language.

	Disagree	Unsure	Agree
Arabic	0	0	100 (7)
Mandarin	0	0	100 (8)

Preferences

You like to learn English by studying grammar.

	Disagree	Unsure	Agree
Arabic	14.3 (1)	14.3 (1)	71.4 (5)
Mandarin	37.5 (3)	12.5 (1)	50.0 (4)

You like to learn English by talking to native speakers.

	Disagree	Unsure	Agree
Arabic	0	0	100 (7)
Mandarin	0	0	100 (8)

You like to learn English by reading magazines and books.

	Disagree	Unsure	Agree
Arabic	14.3 (1)	0	85.7 (6)
Mandarin	0	0	100 (8)

You like to learn English by writing; for example, writing essays, stories, using email.

	Disagree	Unsure	Agree
Arabic	0	14.3 (1)	85.7 (6)
Mandarin	12.5 (1)	0	87.5 (7)

Note. Participants were asked to indicate the response that best described themselves on a scale of "strongly disagree" (1) to "strongly agree" (5). Due to the lack of extreme scores, scores of 1 and 2 were combined into the "disagree" category, while scores of 4 and 5 were combined into the "agree" category.

reported that they liked to learn English by studying grammar. The rest of the participants were either undecided or reported that they did not like to learn English by studying grammar. All of the participants indicated that they liked to learn English by talking to native speakers of English. All but 1 of the participants reported that they liked to learn English by reading: 1 Arabic speaker reported that she did not like to learn English by reading magazines and books. Although 1 Mandarin speaker reported that he did not like to learn English by writing, the other 7 indicated that they liked to learn English by writing. Six of the Arabic speakers responded that they liked to learn English by writing, whereas 1 Arabic speaker was undecided.

Bottom-up reading strategies. Table 6 shows the participants' perceived bottom-up strategy use while answering the CLBA reading questions and when reading silently in English. All of the participants except for 1 Arabic and 1 Mandarin speaker reported that they broke words into smaller parts to help them understand word meanings. All of the Arabic speakers and 5 of the Mandarin speakers indicated that they tried to find synonyms for words in the text, whereas 2 of the Mandarin speakers reported that they were unsure as to whether they used this strategy and 1 Mandarin speaker reported that he did not use this strategy. Five Arabic and 3 Mandarin speakers indicated that they tried to understand the meaning of each word, whereas 2 Arabic and 5 Mandarin speakers reported that they did not focus on the meaning of each word. All but 1 of the Arabic speakers indicated that they focused on the details of the content when reading, whereas of the Mandarin speakers, 5 reported that they focused on the details, 1 was unsure, and 2 indicated that they did not focus on the details of the content. Two of the Arabic speakers and 5 of the Mandarin speakers responded that they focused on the grammatical

Table 6

Percentage of Self-reported Bottom-up Reading Strategies by Language Group (n): Arabic Speakers (n = 7), Mandarin Speakers (n = 8)

When answering the CLBA reading comprehension questions:

you broke words into smaller parts to help you understand their meaning – strategy B1 (breaking lexical items into smaller parts.

	Disagree	Unsure	Agree
Arabic	0	14.3 (1)	85.7 (6)
Mandarin	12.5 (1)	0	87.5 (7)

you tried to find synonyms for words in the text – strategy B3 (identifying synonyms or paraphrases).

	Disagree	Unsure	Agree
Arabic	0	0	100 (7)
Mandarin	12.5 (1)	25 (2)	62.5 (5)

you focused on understanding the meaning of each word.

	Disagree	Unsure	Agree
Arabic	28.6 (2)	0	71.4 (5)
Mandarin	62.5 (5)	0	37.5 (3)

you focused on the details of the content – strategy B2 (scanning for details).

	Disagree	Unsure	Agree
Arabic	14.3 (1)	0	85.7 (6)
Mandarin	25.0 (2)	12.5 (1)	62.5 (5)

you focused on the grammatical structures – strategy B6 (using knowledge of grammar or punctuation).

	Disagree	Unsure	Agree
Arabic	57.1 (4)	14.3 (1)	28.6 (2)
Mandarin	37.5 (3)	0	62.5 (5)

When reading silently in English:

if you don't understand something, you look up unknown words in a dictionary.

	Disagree	Unsure	Agree
Arabic	42.9 (3)	0	57.1 (4)
Mandarin	37.5 (3)	0	62.5 (5)

you focus on mentally sounding out parts of the words.

	Disagree	Unsure	Agree
Arabic	0	14.3 (1)	85.7 (6)
Mandarin	12.5 (1)	12.5 (1)	75.0 (6)

you focus on being able to pronounce each word silently to yourself.

	Disagree	Unsure	Agree
Arabic	0	14.3 (1)	85.7 (6)
Mandarin	62.5 (5)	0	37.5 (3)

Note. Participants were asked to indicate the response that best described themselves on a scale of "strongly disagree" (1) to "strongly agree" (5). Due to the lack of extreme scores, scores of 1 and 2 were combined into the "disagree" category, while scores of 4 and 5 were combined into the "agree" category.

structures when reading and answering the CLBA reading questions. One Arabic speaker was unsure of using this strategy, and 4 Arabic and 3 Mandarin speakers reported that they did not focus on grammatical structures when reading. Four Arabic and 5 Mandarin speakers indicated that when reading silently in English, they look up unknown words, whereas 3 Arabic and 3 Mandarin speakers reported that they do not use this strategy. Six Arabic speakers reported that they mentally sound out parts of words or pronounce each word silently to themselves, whereas 1 Arabic speaker was unsure as to whether he used these two strategies. Six of the Mandarin speakers indicated that they mentally sound out parts of the words when reading silently in English. However, 1 indicated that she was unsure of using this strategy and 1 reported that he did not use this strategy. Five of the Mandarin speakers responded that they do not pronounce each word silently to themselves when reading silently in English and 3 reported that they do pronounce each word silently to themselves.

Top-down reading strategies. Table 7 shows the participants' perceived top-down strategy use while answering the CLBA reading questions and when reading silently in English. Three of the Arabic and 7 of the Mandarin speakers indicated that they recognized the difference between the main points and supporting details when reading and answering the CLBA questions. However, 4 Arabic speakers were unsure as to whether they used this strategy and 1 Mandarin speaker indicated that he did not use this strategy. Six Arabic and 7 Mandarin speakers reported that they related information which comes next in the text to previous information in the text, whereas 1 Arabic speaker was unsure of using this strategy and 1 Mandarin speaker indicated that she did not use this strategy. Six Arabic and 7 Mandarin speakers responded that they related

Table 7

Percentage of Self-reported Top-down Reading Strategies by Language Group (*n*):

Arabic Speakers (*n* = 7), Mandarin Speakers (*n* = 8)

When answering the CLBA reading comprehension questions:

you were able to recognize the difference between the main points and supporting details – strategy T1 (skimming for gist).

	Disagree	Unsure	Agree
Arabic	0	57.1 (4)	42.9 (3)
Mandarin	12.5 (1)	0	87.5 (7)

you were able to relate information which comes next in the text to previous information in the text – strategy T2 (connecting or relating information presented in different parts of the text).

	Disagree	Unsure	Agree
Arabic	0	14.3 (1)	85.7 (6)
Mandarin	12.5 (1)	0	87.5 (7)

you focused on relating the text to what you already knew about the topic – strategy T4 (using background knowledge to speculate beyond the text).

	Disagree	Unsure	Agree
Arabic	0	14.3 (1)	85.7 (6)
Mandarin	0	12.5 (1)	87.5 (7)

you focused on the organization of the text – strategy T5 (recognizing discourse format).

	Disagree	Unsure	Agree
Arabic	28.6 (2)	0	71.4 (5)
Mandarin	50.0 (4)	0	50.0 (4)

Note. Participants were asked to indicate the response that best described themselves on a scale of "strongly disagree" (1) to "strongly agree" (5). Due to the lack of extreme scores, scores of 1 and 2 were combined into the "disagree" category, while scores of 4 and 5 were combined into the "agree" category.

information in the text to what they knew about the topic, whereas 1 Arabic and 1 Mandarin speaker were undecided as to whether they used this strategy. Finally, 5 of the Arabic speakers and 4 of the Mandarin speakers responded that they focused on the organization of the text when reading and answering the CLBA questions, whereas 2 Arabic and 4 Mandarin speakers indicated that they did not use this strategy.

Interrater Agreement

Interrater agreement was assessed to ensure the verbal protocols were coded consistently. Consistency was defined as the extent to which the protocol segments were coded using the same categories by both raters. Consistency is related to the validity of the coding scheme (i.e., its ability to capture the students' behaviours). Of the 456 segments coded, 413 agreements occurred. Therefore, the percentage of total agreement between the researcher and the rater was 90.6%, indicating that the reading strategy segments were consistently coded.

Disagreements, however, were not evenly distributed across the bottom-up and top-down strategy categories: 17 disagreements occurred within the bottom-up strategy categories and 26 within the top-down strategy categories. The disagreements that occurred more than once included the following: the researcher coded seven segments as B5 (matching key vocabulary in the text to key vocabulary in the item) that the rater coded as B3 (identifying synonyms or paraphrases); the researcher coded seven segments as T4 (using background knowledge to speculate beyond the text) that the rater coded as T3 (drawing an inference based on information presented in the text); the researcher coded four segments as T1 (skimming for gist) that the rater coded as T2 (connecting or relating information presented in different parts of the text); the researcher coded four

segments as T3 (drawing an inference based on information presented in the text) that the rater coded as T2 (connecting or relating information presented in different parts of the text); the researcher coded three segments as T4 (using background knowledge to speculate beyond the text) that the rater coded as T2 (connecting or relating information presented in different parts of the text); and the researcher coded two segments as T1 (skimming for gist) that the rater coded as T3 (drawing an inference based on information presented in the text). Therefore, it appeared that the bottom-up strategies were coded with more consistency than the top-down strategies. However, there was no disagreement regarding whether the bottom-up segments reflected bottom-up strategies or whether the top-down segments reflected top-down strategies as all of disagreements were coded with either another bottom-up strategy or another top-down strategy.

Coding Schema

The verbal report procedures 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 data is presented in Table 8. Analyses of the protocols identified 12 main reading strategy categories consistent with those identified in the literature. These included 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

Table 8

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).

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 the strategies are presented in Table 8.

Frequencies of Bottom-up and Top-down Reading Strategies in the Verbal Reports

The total number of bottom-up and top-down strategies assigned to all of the protocol segments for each participant is presented in Table 9. To facilitate the comparison of bottom-up and top-down reading strategies by language group and across participants, strategy frequencies were converted to proportions (i.e., the total number of times a given strategy was reported for each participant was divided by the sum of the total number of strategies reported for each participant). An individual's total proportion score sums to 1.0.

An examination of Table 9 provided insights into the strategies used by the Arabic and Mandarin speakers as they were taking the CLBA Reading Assessment. When reading and answering the CLBA items, both language groups relied most heavily on strategy B3 (finds/identifies synonyms or paraphrases): Arabic speakers used this strategy 24.9% of the time, while Mandarin speakers used it 29.1% of the time. Both groups also used each of the following strategies more than 5.0% of the time: B2 - scanning for details, B5 - matching key vocabulary in the text to key vocabulary in the item, T3 - drawing an inference based on information presented in the text, and T4 - using background knowledge to speculate beyond the text. The most frequently used strategies in descending order for the Arabic speakers were B3 (24.9%), B5 (22.8%), T4 (17.9%), B2 (12.5%), T3 (9.5%), and T2 (5.0%). The most frequently used strategies in descending order for the Mandarin speakers were B3 (29.1%), B5 (24.3%), B2 (12.9%),

Table 9

Bottom-up, Top-down Strategy Frequencies (f) and Proportion (Prop) Scores for Each Participant and Language Group on the CLBA Reading Assessment

Strategy	B1		B2		B3		B4		B5		B6		B7		T1		T2		T3		T4		T5		Total
Participant	*f*	*Prop*	*f*	*Prop*	*f*	*Prop*	*f*	*Prop*	*f*	*Prop*	*f*	*Prop*	*f*	*Prop*	*f*	*Prop*	*f*	*Prop*	*f*	*Prop*	*f*	*Prop*	*f*	*Prop*	*f*
Arabic 1	0	0.000	14	0.136	21	0.204	0	0.000	15	0.146	0	0.000	0	0.000	9	0.087	4	0.039	9	0.087	29	0.282	2	0.019	103
Arabic 2	1	0.011	10	0.114	28	0.318	2	0.023	34	0.386	0	0.000	0	0.000	1	0.011	3	0.034	6	0.068	1	0.011	2	0.023	88
Arabic 3	0	0.000	9	0.120	16	0.213	0	0.000	8	0.107	0	0.000	0	0.000	1	0.013	6	0.080	10	0.133	24	0.320	1	0.013	75
Arabic 4	0	0.000	7	0.092	20	0.263	0	0.000	17	0.224	0	0.000	0	0.000	4	0.053	6	0.079	5	0.066	15	0.197	2	0.026	76
Arabic 5	0	0.000	8	0.129	13	0.210	0	0.000	16	0.258	0	0.000	1	0.016	2	0.032	5	0.081	7	0.113	9	0.145	1	0.016	62
Arabic 6	0	0.000	12	0.174	18	0.261	0	0.000	19	0.275	0	0.000	1	0.014	2	0.029	4	0.058	9	0.130	4	0.058	0	0.000	69
Arabic 7	0	0.000	10	0.118	23	0.271	1	0.012	18	0.212	0	0.000	0	0.000	4	0.047	2	0.024	7	0.082	18	0.212	2	0.024	85
Total/Mean	1	0.002	70	0.125	139	0.249	3	0.005	127	0.228	0	0.000	2	0.000	23	0.041	30	0.054	53	0.095	100	0.179	10	0.018	558
Mandarin 1	0	0.000	12	0.136	18	0.205	1	0.011	28	0.318	0	0.000	2	0.023	3	0.034	6	0.068	10	0.114	6	0.068	2	0.023	88
Mandarin 2	2	0.023	12	0.136	24	0.273	0	0.000	15	0.170	0	0.000	5	0.057	2	0.023	6	0.068	10	0.114	10	0.114	2	0.023	88
Mandarin 3	1	0.012	10	0.122	22	0.268	2	0.024	16	0.195	0	0.000	2	0.024	2	0.024	5	0.061	11	0.134	8	0.098	3	0.037	82
Mandarin 4	1	0.008	13	0.102	55	0.430	2	0.016	20	0.156	3	0.023	1	0.008	3	0.023	9	0.070	5	0.039	13	0.102	3	0.023	128
Mandarin 5	0	0.000	15	0.149	23	0.228	0	0.000	30	0.297	2	0.020	2	0.020	7	0.069	1	0.010	5	0.050	16	0.158	0	0.000	101
Mandarin 6	0	0.000	12	0.121	28	0.283	2	0.020	29	0.293	1	0.010	1	0.010	3	0.030	1	0.010	11	0.111	7	0.071	4	0.040	99
Mandarin 7	0	0.000	10	0.093	29	0.271	2	0.019	36	0.336	0	0.000	1	0.009	2	0.019	6	0.056	14	0.131	3	0.028	4	0.037	107
Mandarin 8	0	0.000	14	0.203	23	0.333	2	0.029	11	0.159	0	0.000	2	0.029	4	0.058	3	0.043	6	0.087	1	0.014	3	0.043	69
Total/Mean	4	0.005	98	0.129	222	0.291	11	0.014	185	0.243	6	0.008	16	0.021	26	0.034	37	0.049	72	0.094	64	0.084	21	0.028	762

Note. Bottom-up strategies: 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 interpret a word or phrase. Top-down strategies: T1 - skimming for gist, T2 - connecting or relating information 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.

T3 (9.4%), T4 (8.4%), and T2 (4.9%).

An examination of the information in Table 9 on a case-by case basis indicated that the first Arabic speaking participant's verbal report (which was given mainly in English) contained a total of 103 bottom-up and top-down strategy segments. This total was the highest number of strategies recorded for any of the Arabic speakers and was spread over eight different strategy categories; this participant did not use four of the strategies (i.e., B1 - breaking lexical items into smaller parts, B4 - matching key words to key visuals, B6 - using knowledge of grammar or punctuation, and B7 - using local context cues interpret a word or phrase). The strategies this participant used at least 5.0% or more of the time when reading and answering the CLBA Reading Assessment items included B2 - scanning for details, B3 - identifying synonyms or paraphrases, B5 - matching key vocabulary in the text to key vocabulary in the item, T1 - skimming for gist, T3 - drawing an inference based on information presented in the text, and T4 - using background knowledge to speculate beyond the text. The strategy this reader used to the greatest extent was T4 - using background knowledge to speculate beyond the text (used 28.2% of the time). Interestingly, despite the heavy strategy use, this participant was the second lowest scoring Arabic speaking participant on the CLBA Reading Assessment, 10/32 = 31.3% (see Table 10 for a comparison of the participants' CLBA scores and the total proportions of bottom-up and top-down strategies used).

The fourth Mandarin speaking participant's verbal report (which was given mainly in Mandarin) contained a total of 128 strategy segments. This total was the highest number of strategies recorded for any of the Mandarin speakers and was spread over all 12 strategy categories. The strategies this participant used at least 5.0% or more of the time included B2 - scanning for details, B3 - identifying synonyms or

Table 10

Verbal Report Participants' CLBA Reading Scores Compared with the Proportion of Total Bottom-up and Top-down Strategies Used

Participant	CLBA Score /32	Total Bottom-up	Total Top-down
Arabic 1	10	0.485	0.515
Arabic 2	16	0.852	0.148
Arabic 3	15	0.440	0.560
Arabic 4	8	0.579	0.421
Arabic 5	19	0.613	0.387
Arabic 6	15	0.725	0.275
Arabic 7	11	0.612	0.388
Arabic *M*	13.43	0.613	0.387
Arabic *Mdn*	15.0		
Arabic *SD*	3.87		
Mandarin 1	23	0.693	0.307
Mandarin 2	21	0.659	0.341
Mandarin 3	23	0.646	0.354
Mandarin 4	22	0.742	0.258
Mandarin 5	16	0.713	0.287
Mandarin 6	26	0.737	0.263
Mandarin 7	26	0.729	0.271
Mandarin 8	16	0.754	0.246
Mandarin *M*	21.63	0.711	0.289
Mandarin *Mdn*	22.5		
Mandarin *SD*	3.89		

paraphrases, B5 - matching key vocabulary in the text to key vocabulary in the item, T2 - connecting or relating information presented in different parts of the text, and T4 - using background knowledge to speculate beyond the text. The strategy this reader used to the greatest extent was B3 - identifying synonyms or paraphrases (used 43.0% of the time). This participant scored near the Mandarin participants' mean on the CLBA Reading Assessment (22/32 = 68.8%).

Although Arabic speaker 5 had the lowest number of strategies recorded for any of the Arab participants, her score on the CLBA Reading Assessment (19/32 = 59.4%)

was the highest in the Arab group. In contrast, the Mandarin speaker (8) with the lowest number of strategies (n = 68) recorded for any of the Chinese participants received the lowest CLBA reading score in the Chinese group (16/32 = 50.0%). However, the other Chinese participant (5) who also scored 50.0% on the CLBA Reading Assessment had the third highest number of strategies (n = 101) in the Chinese group.

A comparison of the strategies employed by the two groups in Table 9 indicated that both groups used similar proportions of strategies in several of the strategy categories when reading and answering the CLBA questions (e.g., B2 almost 13.0% of the time; T2 almost 5.0% of the time; and T3 almost 9.0% of the time). However, Table 10 shows the Mandarin speakers generally used a greater proportion of bottom-up strategies than the Arabic speakers (Mandarin 71.1% vs. Arabic 61.3%), and the Arabic speakers generally used a greater proportion of top-down strategies than the Mandarin speakers (Arabic 38.7% vs. Mandarin 28.9%). More specifically, the Arabic speakers used T4 - using background knowledge to speculate beyond the text - approximately 10.0% more often than the Mandarin speakers (see Table 9). In contrast, the Mandarin speakers used B3 - identifying synonyms or paraphrases - approximately 4.0% more often than the Arabic speakers.

Arabic speaker 2's inferred strategy use, however, was contrary to the general Arab tendencies described above. She used the highest proportion of bottom-up strategies (85.2%) and the lowest proportion of top-down strategies (14.8%) recorded for any of the participants in either of the groups (see Table 10). Interestingly, she was the youngest Arabic speaker in the group with the highest number of years of studying EFL (i.e., 15 years), and the second highest score of the Arab participants on the CLBA Reading

Assessment (16/32 = 50.0%). In addition, she was multilingual as she had also studied German in high school.

With the exception of Arabic speaker 2, Arabic speaker 6 also used a greater proportion of bottom-up strategies than the other Arab participants. Consequently, his pattern of strategy use appeared to be quite similar to the Mandarin pattern of strategy use. Although Mandarin speaker 5 used T4 to a greater extent than the other Chinese participants, her overall pattern of strategy use was in keeping with the other Chinese participants' usage.

Part 2 - Differential Item and Bundle Functioning Results

Sample

Demographic information on the Arabic and Mandarin speakers in Part 2 of the study is provided in Table 11. The Arabic speaking sample consisted of 167 males and 83 females. The mean age of the Arabic sample was 33.55. All of the Arabic speaking immigrants had between 11 and 21 years of education (M = 15.81). Their mean length of residence in Canada at the time of testing was 7 months, 19 days. The Arabic speaking immigrants were from 14 different countries located throughout the Middle East and Northern Africa including Algeria (n = 18), Egypt (n = 40), Iran (n = 1), Iraq (n = 62), Jordan (n = Jordan), Kuwait (n = 8), Lebanon (n = 27), Libya (n = 12), Morocco (n = 12), Saudi Arabia (n = 2), Sudan (n = 34), Syria (n = 21), Tunisia (n = 4), and Yemen (n = 2). In addition to being fluent in oral and written Arabic, 51 of the Arabic speaking immigrants could read, write, and speak a second language: French (n = 39), Dinka (n = 2), Farsi (n = 2), Italian (n = 2), Assyrian (n = 2), Malay (n = 1), Portuguese (n = 1), Turkish (n = 1), and Spanish (n = 1).

Table 11

Demographic Information from Part 2 of the Study

Demographic Variables	Arabic $n = 250$			Mandarin $n = 250$		
	M	*SD*	*Range*	*M*	*SD*	*Range*
Age (years)	33.55	7.39	17-56	33.98	4.99	23-52
Education (years)	15.81	2.29	11-21	15.44	1.66	11-22
Length of residence in Canada (months)	7.63	7.20	1-24	7.53	8.23	1-24

The Mandarin speaking sample in Part 2 of the study consisted of 107 male and 143 female immigrants, all from mainland China. The mean age of this group was 33.98. Their years of education ranged from 15 to 23 years ($M = 15.44$). At the time of testing, the mean length of residence in Canada for the Mandarin speakers was 7 months, 16 days. Two of the Mandarin speakers reported that they could speak and write Japanese, while 3 indicated that they spoke Cantonese, and 1 reported that he spoke Spanish.

A comparison of the demographic information for the participants in Parts 1 and 2 of the study (see Table 12) revealed that the Arabic and Mandarin speaking participants in Part 1 of the study had lived in Canada longer than those in the Part 2 sample at the time of testing. However, the mean age and years of education for both groups of Arabic and Mandarin speakers were quite similar. Therefore, it was concluded that the VPA participants in Part 1 of the study were relatively comparable to those in the Part 2 sample.

Table 12

Comparing Demographic Information from Parts 1 and 2 of the Study

Demographic Variables	Part 1 Arabic $n = 7$			Part 2 Arabic $n = 250$			Part 1 Mandarin $n = 8$			Part 2 Mandarin $n = 250$		
	M	*SD*	*Range*	*M*	*SD*	*Range*	*M*	*SD*	*Range*	*M*	*SD*	*Range*
Length of residence in Canada (months)	16.29	7.54	7-24	7.63	7.20	1-24	13.43	7.57	6-24	7.53	8.23	1-24
Age (years)	32.86	7.71	21-45	33.55	7.39	17-56	36.13	7.40	25-48	33.98	4.99	23-52
Education (years)	14.86	2.41	12-19	15.81	2.29	11-21	16.94	2.76	15-23	15.44	1.66	11-22

Item Coding

The bottom-up, top-down reading strategy classification schema developed in Part 1 of the study provided the conceptual framework for classifying the CLBA reading items into bundles that reflected 12 main reading strategies: 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 interpret a word or phrase, T1 - skimming for gist, T2 - connecting or relating information 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. These 12 strategies served as the organizing principle that was used to group the items into categories for the differential item and bundle functioning analyses described below.

Three item reviewers coded the 32 CLBA reading items using the coding schema presented above. They were asked to code each item according to the reading strategy that they believed was most instrumental in arriving at the correct answer (i.e., each item was classified according to the "most salient" strategy believed to be elicited by the item). All three raters had experience teaching ESL learners to read. Two of the raters were university professors with extensive ESL teaching and teacher-training experience, while the other rater was the researcher, who had completed a Masters in TESL. After the reviewers had independently coded the items, coding consistency was checked by computing the percentage of agreement among the three raters.

Table 13 contains a summary of the interrater agreement for the three judges. The percentage of agreement for the raters was a follows: 75.0% for raters 1 and 2; 90.6% for raters 1 and 3; and 78.1% for raters 2 and 3. Thus the mean rater agreement was 81.3%. Of the nine items that did not receive 100% agreement, five of the disagreements occurred within the same top-down or bottom-up classifications: two raters coded items C2 and C8 as T3, while the other rater coded these two items as T2; two raters coded item C6 as B5, while the other rater coded it as B3; two raters coded item D1 as B3, while the other rater coded it as B5; and two raters coded item D8 as B3, while the other rater coded it as B2. The other four disagreements occurred across the bottom-up and top-down categories: two raters coded item A3 as T2, while the other rater coded it as B3; two raters coded item A5 as T2 while the other coded it as B2; two raters coded item C4 as T3, while the other rater coded it as B2; and two raters coded item D6 as B3, while the other rater coded it as T4.

For three of the nine items that did not receive 100% agreement (i.e., items A3, C2, and C6), the rater in disagreement gave the strategy category that the other two raters

Table 13

Interrater Agreement and Consensus Codes for Three Expert Item Reviewers

CLBA Reading Item	Number of Agreements for All Judges	Number of Agreements/Number of Judges	Bottom-up or Top-down Consensus Code
Task A Item 1	3	1.00	B1
Task A Item 2	3	1.00	T2
Task A Item 3	2	.67	T2
Task A Item 4	3	1.00	B5
Task A Item 5	2	.67	T2
Task A Item 6	3	1.00	B4
Task B Item 1	3	1.00	T5
Task B Item 2	3	1.00	B2
Task B Item 3	3	1.00	B2
Task B Item 4	3	1.00	B2
Task B Item 5	3	1.00	B2
Task B Item 6	3	1.00	T3
Task B Item 7	3	1.00	B2
Task B Item 8	3	1.00	T4
Task C Item 1	3	1.00	T1
Task C Item 2	2	.67	T3
Task C Item 3	3	1.00	B3
Task C Item 4	2	.67	T3
Task C Item 5	3	1.00	B5
Task C Item 6	2	.67	B5
Task C Item 7	3	1.00	T3
Task C Item 8	2	.67	T3
Task D Item 1	2	.67	B3
Task D Item 2	3	1.00	B3
Task D Item 3	3	1.00	B3
Task D Item 4	3	1.00	B5
Task D Item 5	3	1.00	T3
Task D Item 6	2	.67	B3
Task D Item 7	3	1.00	B3
Task D Item 8	2	.67	B3
Task D Item 9	3	1.00	T3
Task D Item 10	3	1.00	T3

Note. Bottom-up strategies: 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 interpret a word or phrase. Top-down strategies: T1 - skimming for gist, T2 - connecting or relating information 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.

had rated as "most salient" a rating of 4 on the saliency rating scale (i.e., 1 = not at all salient to 4 = very salient). This meant that although the deviant rater had selected a different "most salient" strategy than the other two raters, it was obvious that she agreed that the strategy was very instrumental in answering the item. In general, an examination of the raters' disagreements did not reveal any distinct patterns of disagreement in either the bottom-up or top-down item types (i.e., four items that two of the raters had classified as bottom-up items had one rater disagree with the ratings, and five items that two of the raters had classified as top-down items had one rater disagree with the ratings).

Although the interrater agreement percentages between rater 2 and the other two raters were lower than what is normally taken as evidence that a coding schema is stable, after the raters discussed their individual item codings, the raters reached 100% agreement on all of the codings. During the discussion, it became evident that in some cases it was possible to arrive at the correct answer using strategies that one or more of the raters had not previously considered. This realization led to a discussion regarding (a) which of the strategies the raters had deemed "most salient" would clearly be the most helpful or instrumental in answering the question, and (b) how likely the examinees would be to use the strategies. This focus facilitated the ensuing discussion of the nine items where disagreements occurred. In the end, the fact that the raters reached 100% agreement on all of the item codings indicated that the coding schema was a potentially useful theoretical organizing principle for grouping the CLBA reading items into the bundles that were analyzed in the second part of the study.

The distribution of items across the bottom-up and top-down reading strategy categories based on the expert raters' consensus codes is presented in Table 14. The

Table 14

Distribution of CLBA Reading Items Across the Bottom-Up and Top-Down Reading Strategy Categories Based on the Expert Raters' Consensus Codes

Strategy[1]	Number of Items	Task/Item[2]
Bottom-up		
B1 - Lexical	1	A1
B2 - Scanning	5	B2 B3 B4 B5 B7
B3 - Synonym/Paraphrase	7	C3 D1 D2 D3 D6 D7 D8
B4 - Visuals	1	A6
B5 - Matching key words	4	A4 C5 C6 D4
B6 - Syntax	0	
B7 - Local context	0	
Top-down		
T1 - Skimming	1	C1
T2 - Connecting	3	A2 A3 A5
T3 - Inferencing	8	B6 C2 C4 C7 C8 D5 D9 D10
T4 - Speculating	1	B8
T5 - Discourse Format	1	B1

Note. [1]Bottom-up strategies: 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 interpret a word or phrase. Top-down strategies: T1 - skimming for gist, T2 - connecting or relating information 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.
[2]Item numbers: the letters represent the CLBA Tasks A-D and the numbers represent the items corresponding to each passage (e.g., A1 means Task A, item 1).

number of items was unevenly distributed across the bottom-up and top-down strategy categories: 18 items were classified as bottom-up strategy items, and 14 items were classified as top-down items. The strategy category with the highest number of CLBA items was the top-down strategy, T3 - drawing an inference based on information presented in the text (25.0% of the items), followed by the bottom-up strategies, B3 - identifying synonyms or paraphrases (21.9%), and B2 - scanning for details (15.6%). When the raters identified the most instrumental strategy for answering each item, they did not use two of the coding categories (B6 - using knowledge of grammar or punctuation, and B7 - using local context cues interpret a word or phrase). Therefore, these two strategy categories were not tested in the DBF analyses.

Multiple item codes. Although the statistical procedure used in Part 2 of the study (i.e., SIBTEST) is based on the assumption that each item can be classified into one "most salient" strategy category, test items often elicit more than one salient strategy (Gierl, Bisanz, Bisanz, & Boughton, 2003). Table 15 contains the strategies that were rated as salient for each of the CLBA reading items by at least two of the three item reviewers. A salient strategy received a saliency code of 3 = salient or 4 = very salient (see Appendix J). The items in Table 15 are organized according to the reviewers' bottom-up and top-down consensus codes. The mean number of salient categories identified by the item reviewers per item was 1.91. The most frequently used categories were B5 - matching key vocabulary in the text to key vocabulary in the item (identified as salient in 16 or 50.0% of the items) and B3 - identifying synonyms or paraphrases (identified as salient in 11 or 34.4% of the items).

Table 15

All Salient Strategies Identified as Useful by the Item Reviewers When Answering Each CLBA Reading Item

Task/Item	Expert Raters' Item Consensus Code	"Salient" Strategies[1]
	Bottom-up	
A1	B1 - Lexical	B1 B7 T2
B2	B2 - Scanning	B2 B5 T4
B3	B2 - Scanning	B2 B5
B4	B2 - Scanning	B2 B5
B5	B2 - Scanning	B2 B5
B7	B2 - Scanning	B2 B5
C3	B3 - Synonym/Paraphrase	B3
D1	B3 - Synonym/Paraphrase	B3 B5
D2	B3 - Synonym/Paraphrase	B2 B3 T4
D3	B3 - Synonym/Paraphrase	B2 B3
D6	B3 - Synonym/Paraphrase	B3
D7	B3 - Synonym/Paraphrase	B3 B5
D8	B3 - Synonym/Paraphrase	B3 B5
A6	B4 - Visuals	B4
A4	B5 - Matching key words	B2 B5 T4
C5	B5 - Matching key words	B5
C6	B5 - Matching key words	B2 B3 B5
D4	B5 - Matching key words	B5
	Top-down	
C1	T1 - Skimming	T1
A2	T2 - Connecting	B5 T2
A3	T2 - Connecting	B3 T2
A5	T2 - Connecting	B5 T2 T3
B5	T3 - Inferencing	B3 T3
C2	T3 - Inferencing	T2 T3
C4	T3 - Inferencing	B2 B5 T3
C7	T3 - Inferencing	T3
C8	T3 - Inferencing	B5 T2 T3
D5	T3 - Inferencing	T2 T3
D9	T3 - Inferencing	T3
D10	T3 - Inferencing	B3 T3
B8	T4 - Speculating	T4
B1	T5 - Discourse Format	B2 T5

Note. [1]A strategy was considered "salient" if at least two of the item reviewers gave it a saliency code of 3 = salient or 4= very salient, as shown in Appendix J. Saliency was defined in terms of how important or useful the strategy was in answering the item and how likely the examinees were to use the strategy.

Psychometric Characteristics of the CLBA Reading Assessment

Table 16 contains the psychometric characteristics for the CLBA Reading Assessment. The mean total test scores demonstrated 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 in Table 16 also suggested that the skewness and kurtosis values were similar indicating that the shapes of the distributions were similar for both groups of examinees. Furthermore, the results indicated that overall item difficulty, discrimination, and internal consistency were comparable across groups. Mean item discrimination was calculated by transforming the point biserial correlations for each group using Fisher's Z-transformation, summing the transformed correlations, dividing by the total number of items (i.e., 32), and calculating the antilog to convert back to the mean point biserial correlation.

Single-Item SIBTEST Results

Since the raters 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 analysis. The $\hat{\beta}_{UNI}$ estimates for the 32 CLBA items grouped into the 10 remaining reading strategy categories are presented in Table 17. The single-item SIBTEST results found 17 of the 32 items to exhibit moderate to high DIF, $|\hat{\beta}_{UNI}| \geq 0.059$ (14 of the 17 differences were significant, $p < .05$). In total, six of the items exhibited moderate DIF ($0.059 \leq \hat{\beta}_{UNI} < 0.088$), and 11 items exhibited large DIF ($\hat{\beta}_{UNI} \geq 0.088$). Of the six moderate DIF items, five items favoured the Arabic speakers and one item favoured the Mandarin

speakers. Of the 11 high DIF items, four items favoured the Arabic speakers and seven items favoured the Mandarin speakers.

Table 16

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

Characteristic	Arabic	Mandarin
Number of Examinees	250	250
Number of Items	32	32
Mean	17.20	18.29
Standard Deviation	5.87	6.20
Kurtosis	-.72	-.37
Skewness	.03	-.02
Mean Item Difficulty	.54	.57
SD Item Difficulty	.21	.20
Mean Item Discrimination[a]	.40	.42
SD Item Discrimination	.12	.11
Internal Consistency[b]	.83	.85

[a]Point Biserial Correlation

[b]Cronbach's alpha coefficient

Table 17

CLBA Single-item SIBTEST Results

B1 item	$\hat{\beta}_{UNI}$	B2 items	$\hat{\beta}_{UNI}$	B3 items	$\hat{\beta}_{UNI}$	B4 item	$\hat{\beta}_{UNI}$	B5 items	$\hat{\beta}_{UNI}$	T1 item	$\hat{\beta}_{UNI}$	T2 items	$\hat{\beta}_{UNI}$	T3 items	$\hat{\beta}_{UNI}$	T4 item	$\hat{\beta}_{UNI}$	T5 item	$\hat{\beta}_{UNI}$
A1	0.106*	B2	-0.089*	C3	0.121*	A6	-0.077	A4	0.011	C1	-0.149*	A2	-0.264*	B6	0.027	B8	-0.054	B1	0.030
		B3	0.300*	D1	0.116*			C5	0.120*			A3	-0.134*	C2	-0.020				
		B4	-0.052	D2	0.068*			C6	0.029			A5	-0.062	C4	-0.072				
		B5	-0.031	D3	0.018			D4	0.176*					C7	-0.026				
		B7	0.047*	D6	0.215*									C8	-0.037				
				D7	-0.012									D5	0.016				
				D8	-0.066									D9	-0.038				
														D10	-0.077*				

Note. [1]Positive $\hat{\beta}_{UNI}$ values favour the Mandarin speakers and negative $\hat{\beta}_{UNI}$ values favour the Arabic speakers.

[2]Bottom-up strategies: 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 and punctuation, B7 - using local context cues interpret a word or phrase. Top-down Strategies: T1 - skimming for gist, T2 - connecting or relating information 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.

[3]Item numbers: the letters represent Tasks A-D and the numbers represent the items corresponding to each passage (e.g., A1 means Task A, item 1).

* $p < .05$

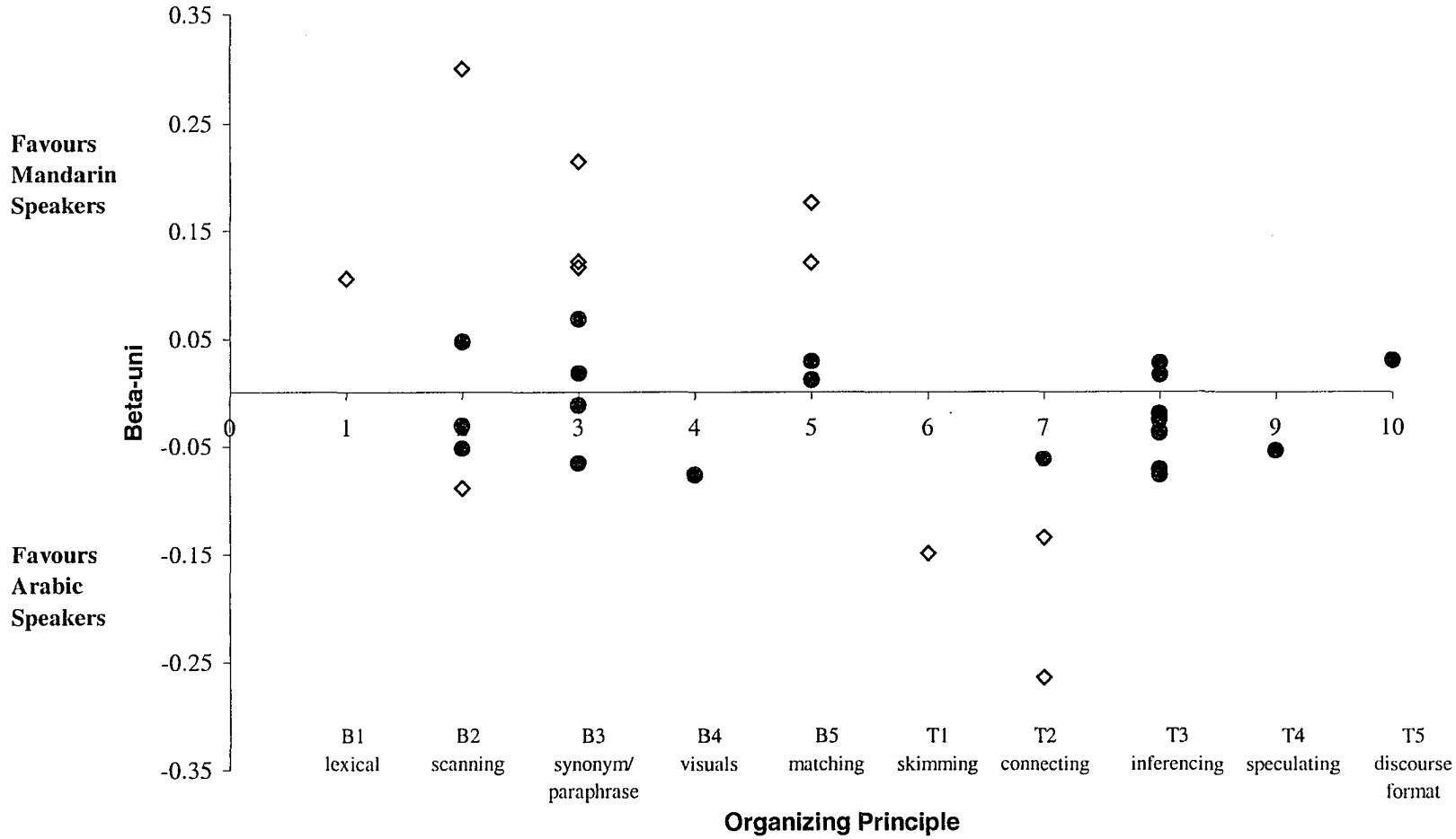
Figure 2 provides a graphical representation of the 32 single-item $\hat{\beta}_{UNI}$ estimates grouped into the bottom-up, top-down reading strategy categories. Positive $\hat{\beta}_{UNI}$ values favour the Mandarin speaking examinees, while negative $\hat{\beta}_{UNI}$ values favour the Arabic speaking examinees. The shaded dots represent moderate and low level DIF items, while the unshaded triangles represent large DIF items.

An inspection of Table 17 and Figure 2 revealed that all of the items in two of the bottom-up strategy categories (i.e., 1 item in B1 and 5 items in B5) and three of the top-down categories (i.e., 1 item in T1, 3 items in T2, and 1 item in T4) were clearly functioning in the predicted manner as the $\hat{\beta}_{UNI}$ estimates in each of these categories in Figure 2 and in the corresponding categories in Table 17 were all in the same direction (i.e., all the $\hat{\beta}_{UNI}$ values in each of these categories were either positive or negative). This indicated that all of the items in two of the bottom-up categories (B1 and B5) favoured the Mandarin speakers and all of the items in three of the top-down categories (T1, T2, and T4) favoured the Arabic speakers. Further examination of Table 17 and Figure 2 indicated that 23 of the 32 CLBA items were functioning in the predicted direction (i.e., 12 of the 18 bottom-up items were found to favour the Mandarin speakers and 11 of the 14 top-down items were found to favour the Arabic speakers).

Differential Bundle Functioning Results

The results for the differential bundle functioning hypothesis tests are reported in Table 18. Each of the bundles was tested against the appropriate matching subtest (i.e., for the bottom-up bundles, the examinees were matched on the 14 top-down items, and for the top-down bundles, the examinees were matched on the 18 bottom-up items).

Figure 2. **Differential Item Functioning Results for the CLBA Reading Assessment**



Note: Each dot or triangle represents the Beta uni ($\hat{\beta}_{UNI}$) estimate for one item. Shaded dots represent small and moderate level A and B DIF items ($|\hat{\beta}_{UNI}| < 0.088$). Unshaded triangles represent high level C DIF items ($|\hat{\beta}_{UNI}| \geq 0.088$).
Positive $\hat{\beta}_{UNI}$ estimates favour the Mandarin speakers and negative $\hat{\beta}_{UNI}$ estimates favour the Arabic speakers.

Seven of the bundles yielded statistically significant $\hat{\beta}_{UNI}$ values ($p < .05$). Four of the bottom-up bundles (B1, B2, B3, B5) significantly favoured the Mandarin speaking examinees over the Arabic speaking examinees, and three of the top-down bundles (T1, T2, T3) significantly favoured the Arabic speaking examinees over the Mandarin speaking examines. However, two of the significant bundles (B1 and T1) consisted of only one item.

Table 18

Differential Bundle Functioning Results

Bundle	Number of CLBA Reading Items	Number of CLBA Items in Matching Subtest	$\hat{\beta}_{UNI}$	Favours
Bottom-up				
B1 - Lexical	1	14	0.176*	Mandarin
B2 - Scanning	5	14	0.233*	Mandarin
B3 - Synonym/Paraphrase	7	14	0.915*	Mandarin
B4 - Visuals	1	14	-0.048	*ns*
B5 - Matching key words	4	14	0.578*	Mandarin
Top-down				
T1 - Skimming	1	18	0.137*	Arabic
T2 - Connecting	3	18	0.463*	Arabic
T3 - Inferencing	8	18	0.445*	Arabic
T4 - Speculating	1	18	0.044	*ns*
T5 - Discourse Format	1	18	-0.033	*ns*

*$p < .05$

CHAPTER V: DISCUSSION AND CONCLUSIONS

This chapter is organized in three sections. In the first section, the research questions and hypotheses are outlined. Brief descriptions of the methods used in Parts 1 and 2 of the study along with a summary and discussion of the key findings are presented in the second section. The final section contains a discussion of the limitations and implications for future research.

Summary of Research Questions and Hypotheses

The purpose of this study was twofold: first, to 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; and second, to identify items/bundles of items that elicit systematic performance differences for the Arabic and Mandarin speaking examinees on Form 1, Stage II of the CLBA Reading Assessment. The following questions were the primary research questions addressed in this study:

1. What are the bottom-up and top-down reading strategies that intermediate proficiency Arabic and Mandarin speaking ESL learners employ when reading and answering the CLBA reading items?
2. Is there evidence for differential item performance for Arabic- and Mandarin-speaking examinees on the CLBA Reading Assessment?
3. If so, is the source of differential performance related to differences in reading strategy use?

More specifically, it was hypothesized that the Arabic speakers would outperform the Mandarin speakers on the five bundles of CLBA reading items that were assumed to

elicit top-down reading strategies, while the Mandarin speakers would outperform the Arabic speakers on the five bundles of items that were assumed to elicit bottom-up reading strategies. Supplementary data were also collected to obtain information on the verbal report participants' CLBA topic familiarity and their passage and item comprehension, perceived reading abilities, preferences for learning English, and perceived strategy use when completing the CLBA Reading Assessment.

Discussion

Question 1: What are the bottom-up and top-down reading strategies that intermediate proficiency Arabic and Mandarin speaking ESL learners employ when reading and answering the CLBA reading test items?

To answer the first question, verbal report data were collected from 7 Arabic and 8 Mandarin speaking intermediate ESL learners as they completed the 32 CLBA reading comprehension items. The verbal reports were conducted to identify, clarify, and elaborate on the reading strategies involved in carrying out the CLBA tasks. Sampling continued until all properties of the reading strategy categories were identified and no new information could be obtained from either language group. This resulted in the bottom-up, top-down reading strategy classification schema that was used to code the CLBA items in the second part of this study (see Table 8).

The coding schema captured definite patterns of similarities and differences among the Arabic and Mandarin speaking participants in the strategies they used when reading and answering the CLBA reading questions. An examination of the reading strategies employed by the two groups of ESL learners in Part 1 of the study (see Tables 10 and 11) indicated that all participants used most of the strategies identified in the

strategy schema. However, the pattern of strategy use in the verbal reports indicated that both groups relied most heavily upon six of the strategies. The most frequently used strategies in descending order for the Arabic speakers were B3 - identifying synonyms or paraphrases (24.9%), B5 - matching key vocabulary in the text to key vocabulary in the item (22.8%), T4 - using background knowledge to speculate beyond the text (17.9%), B2 - scanning for details (12.5%), T3 - drawing an inference based on information presented in the text (9.5%), and T2 - connecting or relating information presented in different parts of the text (5.4%); whereas the most frequently used strategies in descending order for the Mandarin speakers were B3 - identifying synonyms or paraphrases (29.1%), B5 - matching key vocabulary in the text to key vocabulary in the item (24.3%), B2 - scanning for details (12.9%), T3 - drawing an inference based on information presented in the text (9.4%), T4 - using background knowledge to speculate beyond the text (8.4%), and T2 - connecting or relating information presented in different parts of the text (4.9 %).

Both the Arabic and Mandarin speakers relied most heavily upon two of the bottom-up strategies (i.e., B3 - identifying synonyms or paraphrases and B5 - matching key vocabulary in the text to key vocabulary in the item). Anderson et al. (1991) also found that these two strategies were two of the five most frequently employed reading comprehension test-taking strategies by Spanish speaking ESL learners. It is likely that the structure and nature of the CLBA Reading Assessment promotes the use of bottom-up strategies over top-down ones. A substantive analysis of both stages of Form 1 of the CLBA Reading Assessment conducted by the researcher revealed that 49 of the 64 items dealt with lower-level skills that are presumed to elicit bottom-up strategies. This finding

lends support to Hill and Parry (1989, 1992), and Purpura's (1997) conclusion that most ESL reading comprehension test questions require readers to search for specific details or facts, forcing readers to focus on low-level linguistic cues, which tend to elicit bottom-up as opposed to top-down reading strategies.

The verbal report data in this study also revealed that both groups of learners used similar proportions of strategies (within 1.0% of the time) in nine of the 12 strategy categories when reading and answering the CLBA questions (see Table 9). Although there was a fair amount of variation within the two groups, on average, the Mandarin speakers applied a slightly wider range of strategies with greater frequency than the Arabic speakers (i.e., the Mandarin speakers had a tendency to use a wider range of the bottom-up strategies than the Arabic speakers, and the mean number of strategies used by the Mandarin and Arabic speakers was 95 and 80, respectively). Furthermore, as predicted, the Arabic speakers used a greater proportion of top-down strategies than the Mandarin speakers (38.7% vs. 28.9%, respectively). When the Arabic speakers encountered comprehension problems they did not refer closely to the text; rather, they had a greater tendency than the Mandarin speakers to rely upon their own background knowledge or common sense. In contrast, when the Mandarin speakers encountered words or sentences they did not understand, they were more likely to try to analyze them by focusing on low-level linguistic cues.

Since the Arabic speakers in Part 1 of this study appeared to be weaker readers than the Mandarin speakers, as demonstrated by the difference in their mean CLBA reading scores (Arabic 42.0% vs. Mandarin 67.6%), Wolff (1987) and Hammadou (1991) would probably argue that the differences in bottom-up and top-down strategy use were a

function of language proficiency, as they concluded that beginner FL/EFL learners have a tendency to rely upon top-down rather than bottom-up processing strategies. However, their conclusions were based on results from native English speakers studying French or Italian at an American university, and native German speakers studying English in Germany. The similarities between these Indo-European languages and cultures may have influenced the ways in which the language learners approached the L2 texts. In addition, the participants in Part 1 of this study were not novice ESL learners – they were all enrolled in intermediate ESL classes. Therefore, one could argue that the differential strategy use may be attributed to differences in the Arabic and Mandarin speakers' linguistic, cultural, and educational systems. Furthermore, although the Arabic speakers in Part 1 of this study appeared to be weaker readers than the Mandarin speakers, this was not the case in Part 2 of the study as the two groups' mean CLBA reading scores were more comparable (Arabic 53.8%, Mandarin 57.2%). Despite the similarities, the Arabic speakers significantly outperformed equal ability Mandarin speakers on 4 of the 14 CLBA top-down strategy items, while the Mandarin speakers significantly outperformed equal ability Arabic speakers on 9 of the 18 CLBA bottom-up strategy items. This last finding is examined in further detail in the discussion of research Question 3 below.

Preliminary results from this study suggested that successful strategy use is likely a function of linguistic/cultural differences (Parry, 1996). It appeared that ESL learners in these two linguistic/cultural groups had particular reading strategy strengths and weaknesses that may have been related to their experiences with written language and the social process of learning to read (Parry, 1996). It may have been the case that the strategies 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 may 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 tended to encourage the effective use of bottom-up strategies, whereas the exposure the Arab EFL students received to more communicative activities likely promoted the effective use of top-down reading strategies. The results from this study also appeared to support Fender's (2003) results, which implied that Arabic ESL learners would have greater success in using top-down strategies than ESL learners from nonalphabetic L1 backgrounds (e.g., Chinese). It was 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. Finally, the rather low CLBA reading scores attained by the ESL learners in both linguistic/cultural groups demonstrated that immigrants from these language groups would likely benefit from additional instruction in English reading skills and strategies.

Question 2: Is there evidence for differential item performance for Arabic and Mandarin speaking examinees on the CLBA Reading Assessment?

To answer the second question, a statistical analysis of Form 1, Stage II of the CLBA Reading Assessment was conducted using SIBTEST to identify items that functioned differentially for Arabic and Mandarin speaking examinees. The guidelines suggested by Roussos and Stout (1996) were used to classify DIF items in the present

study. Items with moderate or high level ratings, $|\hat{\beta}_{UNI}| \geq 0.059$, were considered DIF items whereas those with negligible ratings, $|\hat{\beta}_{UNI}| < 0.059$, were not considered DIF items. Overall, the moderate and high DIF items were rather evenly distributed between the two groups: nine DIF items were found to favour the Arabic speaking examinees (5 moderate and 4 high DIF items), while eight DIF items were found to favour the Mandarin speaking examinees (1 moderate and 7 high DIF items). When comparing examinees from different L1 backgrounds, finding such a large number of DIF items is not uncommon. For example, for Indo- and non-Indo-European examinees, Ryan and Bachman (1992) found DIF in 65 of 146 items on one form of the TOEFL and 25 of the 40 reading and vocabulary items on one form of the FCE. However, when faced with such large numbers of DIF items researchers usually have difficulty specifying the sources of differential performance (*Standards for Educational and Psychological Testing*, 1999).

Although this study found that 53.1% of the items (i.e., 17 of the 32 items) displayed moderate to large DIF, such item level analyses have generally proved uninterpretable, thereby providing insufficient evidence for making decisions regarding the retention and deletion of test items. If all 17 of these items were removed from the CLBA Reading Assessment, this would have a devastating effect on the original test specifications. Since it is difficult for expert reviewers to substantively interpret such a large number of DIF items without imposing a theoretical framework upon the data, a confirmatory approach to (a) grouping the items based on the bottom-up, top-down ESL reading strategy framework, and (b) statistically testing the groups of items for DBF was used to answer the third primary research question proposed in this study.

Question 3: Is the source of differential performance related to differences in reading strategy use?

To answer the third question, first, a substantive analysis of the CLBA items was conducted by three ESL reading experts to identify items that were presumed to elicit bottom-up or top-down reading strategies. The bottom-up, top-down reading strategy classification schema that was based on a review of the literature and the verbal protocol analyses conducted in Part 1 of the study provided the conceptual framework for classifying the items into bundles that reflected 12 main reading strategies. Therefore, the 12 strategies served as the organizing principle (i.e., the theoretical framework used to group the items) in the differential bundle functioning analyses.

In the item review process, the reading experts were asked to independently classify each of the CLBA reading items according to the reading strategy that they believed was most instrumental in arriving at the answer and also rate the saliency of each strategy for each item (see Appendix J). Once the initial coding was completed, a group discussion and 100% consensus for rating the most salient strategy for each item was conducted.

Next, statistical analyses of the CLBA item level data collected from 250 Arabic and 250 Mandarin speaking immigrants' test forms were conducted using SIBTEST to test bundles of items (classified by the experts using the reading strategy coding schema) that functioned differentially for Arabic and Mandarin speaking examinees. Based on the reading strategy framework, it was predicted that the Mandarin speakers would outperform the Arabic speakers on the five bundles containing items that were assumed to elicit bottom-up strategies, whereas Arabic speakers would outperform Mandarin

speakers on the five bundles of items that were assumed to elicit top-down strategies. These hypotheses were tested using the CLBA Reading Assessment data.

A bundle $\hat{\beta}_{\text{UNI}}$ is interpreted as the expected advantage in number correct score that one group has over the other on the bundle of items. For example, in Table 18, the $\hat{\beta}_{\text{UNI}}$ estimate .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 1 score point more on the seven B3 items than an Arabic speaking examinee of equal ability. In the case that a bundle only has one item, $\hat{\beta}_{\text{UNI}}$ is interpreted as the expected advantage in proportion correct that one group has over the other on the studied item. Therefore, Arabic speaking examinees can be expected to obtain approximately $1/10^{th}$ of a score point more than equal ability Mandarin speaking examinees on the item in bundle T1. 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.

Consistent with expectations based on the reading strategy framework, seven of the hypotheses were supported by the SIBTEST bundle analyses. Systematic group differences were found in four of the bottom-up strategy categories: 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 were found to favour the Mandarin speakers. 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). Since

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 may have contributed to differential bundle functioning (DBF) favouring the Mandarin speakers.

As predicted, systematic group differences were also found for three of the five top-down strategy categories: 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 were found to favour the Arabic speakers. This meant 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). Since 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 may have contributed to differential bundle functioning (DBF) favouring the Arabic speakers. Thus it was concluded that (a) the bottom-up, top-down reading strategy coding schema is a useful framework for explaining the sources of group differences on Form 1, Stage II of the CLBA Reading Assessment, and (b) the results of this study could be used to inform conclusions about bias, equity, and fairness in relation to the CLBA Reading Assessment.

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.

CLBA Reading Assessment bias, equity, and fairness. Evidence of DIF/DBF does not necessarily indicate that individual items or groups of items are biased. What counts as an internal source of bias ultimately depends on the construct definition and how it is operationalized in the specifications of the test content framework. If an item or group of items is measuring a secondary dimension that is an appropriate part of the intended construct, the secondary dimension is considered auxiliary (Roussos & Stout, 1996). Thus the DIF/DBF between groups reflects a true difference in the construct and is considered benign. Alternatively, if an item or a group of items is measuring an unintended secondary dimension, the secondary dimension is considered nuisance as it contains sources of construct-irrelevant difficulty that disadvantage members of subgroups of the population. DIF/DBF caused by nuisance dimensions reflects bias which may 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).

Despite the fact that consensus regarding what constitutes the very essence of reading comprehension is problematic, since bottom-up and top-down strategy reading strategy use may theoretically be considered relevant to the construct being measured by the CLBA Reading Assessment, the DIF/DBF exhibited by the items/bundles may be attributed to an auxiliary dimension of ESL reading comprehension (i.e., bottom-up and top-down processing) and deemed benign. If the content framework of the CLBA Reading Assessment is reflective of the proportion of bottom-up and top-down reading

tasks for the context in which the test results will be interpreted, then the items will be proportionally representative of the target language use (TLU) tasks. However, if the proportion of items in the two strategy categories exceeds the proportion of items deemed to validly represent the construct of ESL reading comprehension in the TLU context, then content relevance and representativeness, and thus test fairness become an issue.

Since ESL placement and proficiency tests cannot adequately reflect the full range of skills that ESL learners employ in real life, experts must be relied upon to determine which skills to assess and which tasks to use to assess these skills. It is the experts'/test developers' underlying values that provide the rationale for prioritizing and selecting what is tested, how it is tested, and why it is tested. Thus, as Madaus (1990) argues, tests should be evaluated for the ways in which they promote specific values and diminish others as the underlying values embedded in the test may have differential effects on examinees from varying socio-cultural, linguistic, and educational backgrounds.

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, however, as no information on the theory of reading or the table of specifications used to shape the assessment was made available to the researcher. 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/DBF in the current study indicates that factors related to bottom-up and top-down processing affect the probability of a correct response. When interpreted in the context of an interactive model of reading, these results

imply that there is a need for a balance between bottom-up and top-down strategy items as equal ability ESL learners from different linguistic/cultural backgrounds are not equally successful when answering bottom-up and top-down reading strategy items. Such a change in the structure of the CLBA Reading Assessment would promote greater equality of opportunity for the examinees as the test would more fairly assess the strengths and weaknesses of the individuals tested. Fair and accurate assessment of immigrants' English language skills is essential as test scores have the power to significantly impact their lives. CLBA results can either provide or take away immigrants' opportunities, most importantly access to instructional programs that meet their needs.

Supplementary Data

The supplementary data collected from the participants in Part 1 of the study revealed that, although the Arabic speakers generally reported being more familiar with the topics of the CLBA reading passages, this did not appear to provide them with any advantage over the Mandarin speakers, as the Arabic speakers' scores were considerably lower than those of the Mandarin speakers (Arabic $M = 13.43$, $Mdn = 15$ vs. Mandarin $M = 21.63$, $Mdn = 22.5$). This result was surprising, as some previous research has suggested that background knowledge has a positive effect on reading comprehension (see Hammadou, 1991; Pritchard, 1990). Thus it appears that self-reports of topic familiarity may not be very accurate indicators of background knowledge. With respect to CLBA passage comprehension, the Mandarin speakers indicated that they understood greater proportions of the reading passages than the Arabic speakers (see Table 3). This finding was reflected in their CLBA reading scores. However, both groups reported that

they had difficulty with the vocabulary in a similar number of items (the Arabic speakers identified 4 items with difficult vocabulary while the Mandarin speakers identified 5 items). Based on the differences in the CLBA group results, one would have expected the Arabic speakers to identify more items with difficult vocabulary than the Mandarin speakers. It is possible that the Arabic speakers were not aware of their lack of success when answering many of the bottom-up vocabulary items because they were approaching the text from a top-down perspective.

Since all of the verbal protocol participants reported that they were good readers in their first language, it was assumed that they did not have any reading disabilities. Not surprisingly, the participants were less confident in their English reading abilities than their L1 reading abilities: only 1 Arabic speaker and 2 Mandarin speakers agreed with the statement "you are a good reader in English." However, these 3 participants' self-reports of their reading abilities were not highly reflective of their CLBA Reading Assessment scores as Arabic speaker 2 scored 50.0%, while Mandarin speakers 3 and 4 scored 71.8% and 68.8%, respectively on the reading subtest (see Table 10). These findings reinforce Kruger and Dunning's (1999) conclusions that participants often have difficulties assessing their own abilities. This is a problem inherent in using self-report data gathered through questionnaires. In addition, these findings indicate that questionnaire items need to clarify terms such as "good," because what is good to one person may not necessarily be good to another.

While most of the participants reported that they liked to learn English by reading, 1 Arabic speaker (Arabic participant 1) indicated that she did not like to read in English because it is very difficult, time consuming, and frustrating for her. Her reaction was not

surprising because in terms of her CLBA reading score, she was the second weakest reader in Part 1 of the study (see Table 10). Interestingly, however, her verbal report data indicated that she used strategies more frequently than any of the other Arabic speakers. This result supports the conclusion that the degree of learner success is not related to the frequency of strategy use, but to the appropriate selection and use of strategies (see Chamot & Kupper, 1989; Vann & Abraham, 1990). If this is the case, it appears that less effective readers need to be taught how and when to use a variety of appropriate strategies when reading in English. As Grabe (2004) suggests, the key to teaching strategies is through a combined-strategies instructional approach rather than as strategies taught independently of one another.

Results of this nature have created an interest in strategy training programs and research to find out whether teachers can help less effective readers improve through strategy instruction, and if so, how reading strategy instruction should be implemented (e.g., Barnett, 1988; Schueller, 2000, in press). Since it is often extremely difficult to get people to change their ways (Argyris, 1970) or create new habits, it is necessary to create an environment where the teachers and students positively embrace strategies and make changes to their beliefs and values regarding strategy use. In other words, teachers and students must be convinced of the value of strategies for them to be perceived as being beneficial. Strategies should not only be perceived useful, they should also be linked to effective performance on various tasks. However, considering the wide linguistic and cultural variation found in most Canadian ESL classes, it may be difficult to convince all students of the value of certain types of strategies as they may resist developing strategies

that have not been emphasized or are not traditionally accepted or valued by the education systems in their countries of origin.

Anderson (1999) provides some valuable suggestions for techniques instructors can use to help learners realize the value of strategies and to develop confidence in using a wide range of reading comprehension strategies that are appropriately matched to different tasks. One technique he recommends for helping students answer reading comprehension questions is to have them go through the process of explaining how they arrived at their answers. Justifying their answers helps the students to develop an awareness of the strategies they use when answering questions (i.e., metacognitive awareness) and often leads them to the realization that they have the wrong answer. Evidence of such self-correction was found in the verbal report data collected in Part 1 of the current study. While reporting retrospectively, several of the students realized they had chosen the wrong answers. These results verify the value of having students monitor their reading comprehension by reflecting on and verbalizing their thinking processes. Several SLA researchers (e.g., Anderson, 2002; Anderson & Vandergrift, 1996; Carrell, 1989; Chamot, 2004; Chamot & O'Malley, 1994) have stressed the need for explicit strategy instruction that encourages learners to monitor and evaluate their strategy use. Anderson (1999) provides several excellent suggestions for teaching learners how to select and apply strategies more effectively through the use of teacher modelling that demonstrates successful orchestration of appropriate strategies and students' verbal reports of their strategy use.

In Part 1 of the current study, although both groups of immigrants appeared to use eight of the reading strategies with similar frequencies, the DIF and DBF results from

Part 2 suggest that success in applying these strategies may be related to linguistic/cultural differences as the Arabic speakers were more successful in answering many of the top-down questions than equal ability Mandarin speakers, and the Mandarin speakers were more successful in answering many of the bottom-up questions than equal ability Arabic speakers. This suggests that the degree of learner success on the CLBA Reading Assessment was not related to the frequency of strategy use, but to the appropriate selection and success in using strategies to select the correct answers (see Chamot & Kupper, 1989; Vann & Abraham, 1990). Pedagogically, these results seem to suggest that the Arabic speakers would benefit from instruction in bottom-up strategies that would not only help them attend to word-level details and the local context of the words, but would also encourage them to evaluate their success in using bottom-up strategies when answering reading comprehension questions. With respect to the Mandarin speakers, they would likely benefit from instruction that helps them learn how to successfully apply appropriate top-down strategies. However, at this initial stage in researching differences in successful and unsuccessful reading strategy use in these two linguistic/cultural groups, these conclusions are tentative.

In the substantive review of the 32 CLBA items conducted in Part 2 of the study, the ESL experts identified 18 bottom-up items and 14 top-down items. The proportion of bottom-up (56.3%) and top-down (43.8%) items corresponded more closely to the proportion of strategies inferred from the Arabic speakers' verbal report data (bottom-up: 61.3 % and top-down: 38.7%) than the Mandarin speakers' verbal report data (bottom-up: 71.1% and top-down: 28.9%) (see Table 10). Although these results suggest that the Arabic speakers were using a greater proportion of top-down strategies than the Mandarin

speakers, this did not necessarily indicate that the Arabic speakers were using the strategies successfully, as at times their background knowledge interfered with their reading comprehension. Therefore, to ensure that the ESL readers who have a tendency to rely on top-down strategies do not over-rely on their background knowledge when answering reading comprehension questions, test developers should emphasize in their instructions that the examinees answer the questions based on the information in the reading passage because reliance on their personal experiences may cause them to choose distracters which may reflect differences in their socio-cultural knowledge and experiences.

Since reading comprehension relies on schematic and systemic knowledge (Widdowson, 1983), ESL instructors often encourage their students to activate their existing knowledge of text topics when reading. However, test tasks are not designed to test facts that readers "ordinarily carry around in their heads" (Hill & Parry, 1992, p. 437). Thus test developers deliberately try to design tasks that cannot be answered without reference to the text even by readers who have specialized knowledge of the text content. An examination of the readers' thought processes revealed in the verbal reports collected in Part 1 of this study indicated that ESL learners would probably benefit from a discussion of reading comprehension test development practices. ESL instructors should emphasize that their students' background knowledge may not be appropriate for the context of the texts being read because the test developer's view of typical background knowledge may drastically differ from their own knowledge (Hill & Parry, 1992). This does not mean that instructors should discourage their students from activating their background knowledge when reading. Rather as Block (1986) suggests,

when responding to reading comprehension questions, readers need to anchor their knowledge-based associations to the information in the text. This would prevent them from merely relying on their background knowledge when answering reading comprehension questions.

A number of preliminary implications for ESL reading theory, and teacher, test, and curriculum development practices can be drawn from the foregoing discussion. First, an interactive model of reading which stresses the importance of both bottom-up and top-down reading strategies appears to be a valid framework that is appropriate for modelling English as second language reading comprehension. 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 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 first language and culture on reading so 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. In addition, 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

Verbal Report and Supplementary Data Limitations and Implications

Although the verbal report data were collected in a low anxiety situation that did not have any real life implications for the participants and therefore did not simulate actual testing conditions, all of the participants were genuinely motivated as they approached the task of answering the questions correctly with effort and persistence. Furthermore, since the classification schema was developed from the strategies that were elicited by the CLBA items, it was not intended to be a comprehensive account of all possible bottom-up and top-down reading strategies. Nevertheless, it reflected the key mental operations the participants in this study used when answering CLBA constructed-response and multiple-choice reading comprehension questions.

A comparison of the self-report strategy questionnaire data (see Tables 8 and 9) with the verbal report data (see Tables 10 and 11) collected in Part 1 of the study, revealed that the students' perceptions of their strategy use were not exactly commensurate with the strategies inferred from the verbal reports. For example, even though most of the participants (i.e., 6 of the Arabic speakers and 7 of the Mandarin speakers) reported that they break words into smaller parts to help them understand the meaning of words (strategy B1), evidence of this strategy was only found in 1 Arabic speaker's and 3 Mandarin speakers' verbal reports. In addition, while 3 Arabic and 7 Mandarin speakers indicated that they were able to recognize the difference between the main points and supporting details (strategy T1) evidence of this strategy was found in all the participants' verbal reports.

Differences in the self-report and verbal report data may have been attributed to five sources. First, the participants may have (a) made guesses about what they actually do when they read. Consequently, they may have over- or under-estimated the frequency of their strategy use when responding to the questionnaire (Cohen & Scott, 1996). Second, because Canadian ESL teachers often encourage their students to use many of the reading strategies identified in the questionnaire, the participants may have distorted their self-reports by over-reporting socially desirable responses and under-reporting negative ones. Third, although bilingual interpreters assisted with the administration of the strategies questionnaire, the participants may still have misunderstood some of the questions. Fourth, the act of having to report verbally while reading and answering the CLBA questions may have disrupted the participants' normal reading behaviours so additional or different strategies were employed (Cohen & Scott, 1996). Finally, the participants' thought processes may not have been accessible or easily verbalized (Cohen & Scott, 1996) despite the fact that they had the option to report in their L1. The differences in the self-report questionnaire and verbal report data reinforce the need to cross check the data by using multiple data collection procedures (i.e., triangulation). Further research comparing the self-perception of reading strategies with inferred strategy use is clearly in order.

As with all case study research, the patterns of strategy use inferred from the verbal report data must be considered as hypotheses to be tested in future studies conducted with larger groups. Thus, another implication for future research that is evident from this research is that the verbal report sample should be increased. Then for the DBF analyses, each CLBA item could be classified according to the most frequently used

reading strategy rather than the strategy that the reading experts believe to be the most salient in answering each question. Increasing the VPA sample would also allow for statistical analyses of group differences in the frequency of strategy use, perceived strategy use, strategy preferences, topic familiarity, and passage and item comprehension. Unfortunately, however, such an increase would be more resource and time intensive.

A further implication for future research that is evident from this study is that investigations of the relationship between the L1 and ESL/EFL reading strategies employed by learners in these two linguistic/cultural groups are necessary to provide information regarding the extent to which learners from these groups use bottom-up and top-down strategies when processing texts in both their L1 and English. Such studies would help to clarify the effects of orthography and culture on differences in strategic processing between Arabic and Mandarin speaking examinees.

Future ESL reading strategy training studies comparing control groups with experimental groups taught bottom-up or top-down reading strategies would also be informative. For example, a study which compares an Arabic speaking control group with Arabic speakers taught to successfully apply bottom-up reading strategies, and a study of Mandarin speaking immigrants which compares a control group with an experimental group taught top-down reading strategies would provide evidence which either supports or refutes the preliminary pedagogical directives outlined above.

Differential Item and Bundle Functioning Limitations and Implications

An examination of the $\hat{\beta}_{\text{UNI}}$ values in Figure 2 showed that there was variation in the way the items were functioning across the bottom-up and top-down categories. For example, the items in the B2, B3, and T3 bundles did not consistently favour one group

over the other, and the items in the B4 and T5 bundles functioned in the opposite direction than predicted. In other words, although the item in B4 was predicted to favour Mandarin speakers, it was found to favour Arabic speakers, and although the item in T5 was predicted to favour Arabic speakers, it was found to favour Mandarin speakers. Thus, the reading strategy framework tended to inconsistently identify group performance differences for items in the following five strategy categories: B2 - scanning for details, B3 - identifying synonyms or paraphrases, B4 - matching words to key visuals, T3 - drawing an inference based on information presented in the text, and T5 - recognizing discourse format. In addition, although the item in T4 was functioning in the predicted direction, the difference between the groups was not significant. The lack of significance, however, may have been a function of the relatively small group sample sizes used in the current study. Finally, since researchers do not agree upon how to distinguish statistical from practical significance in DBF research, future research investigating and developing guidelines for interpreting bundle effect size measures is necessary. This would reduce the over-reliance on statistical significance testing in current DBF analyses (Gierl et al., 2003).

As Douglas et al. (1996) suggested, the DIF occurring among the individual item bundles should be carefully examined to gain a better understanding of additional secondary dimensions and causes of DIF operating within the bundles. Unfortunately, a content analysis of the item(s) in the five bundles with inconsistent DIF patterns (identified in the previous paragraph) did not reveal any explanations for the departures from the predicted DIF patterns. Perhaps the presence of other unidentified construct relevant or irrelevant dimensions (i.e., auxiliary or nuisance dimensions) contributed to

the inconsistencies in the patterns of DIF operating within these bundles. As Bolt (2002) suggested, several nuisance dimensions may work together to impact performance on items within a bundle. For example, distinct item format effects (e.g., constructed-response versus multiple-choice) or passage topic effects might be regarded as additional dimensions operating within the secondary bottom-up, top-down strategy dimensions in the CLBA Reading Assessment. It was also possible that the departures from the predicted DIF/DBF patterns were caused by the vulnerability of the items to misclassification by the expert judges. Therefore, an additional problem that complicated the DBF analyses in this study involved the item classification schema.

Although coding the items using the reading strategy framework was fairly straightforward, at times the coders found it difficult to classify the items into one specific "salient" category. This was not surprising as other researchers have also found it difficult to anticipate the cognitive processes examinees use to answer the questions correctly (Gierl et al., 2001). In addition, the reviewers sometimes identified a combination of both bottom-up and top-down strategies as essential to answering certain questions. This finding is consistent with interactive theories of reading (e.g., Rumelhart, 1977; Stanovich, 1980, 2000), which propose that readers simultaneously or alternately use bottom-up and the top-down strategies to construct meaning. These findings suggested that researchers need to closely analyze the cognitive demands of the CLBA reading items in order to develop a more complex representation of the construct of strategic reading comprehension. Presumably, this would lead to the development of additional organizing principles that might assist in explaining and interpreting group performance differences on the CLBA Reading Assessment.

As Gierl et al. (2003) suggested, a potential solution to the problem of using models where items can only be classified into one category is to use a statistical approach based on multidimensional IRT (e.g., Ackerman, Gierl & Walker, 2003; Bolt, 2002). The cognitive complexity elicited by reading test items could be more accurately modeled by using multidimensional models that allow for the items to be classified into multiple reading strategy categories and thereby model the strategies that readers successively orchestrate to achieve the goal of correctly answering a reading comprehension question.

Although the results of this study suggested that DIF/DBF on the CLBA reading Assessment appears to be associated with reading strategies that may be specific to group membership and first language 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/DBF hypotheses, which may provide further insights into the causes of DIF/DBF (Stout & Roussos, 1995). In addition, since the results of the current study were based on a limited item pool (32 items), follow-up substantive and statistical DIF/DBF 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.

Following Gierl, Bisanz, and Bisanz' (2001) recommendations for developing an interpretative framework for understanding group performance differences, further research is also required to validate the dimensional interpretations and clarify why the group differences occur on the CLBA. This would require much larger sample sizes than those used in the current study. In reality, however, it is difficult to obtain sufficiently large samples when conducting research on many second language tests. Since no large database of CLBA item level data was in existence when the current study was conducted, it took the researcher almost two years to gain permission to conduct this study and collect the CLBA data at the two institutions in Alberta. Given that the CLBA is now a high-stakes test, the primary administrators of the CLBA should establish policies to facilitate multiple forms of research designed to validate the CLBA test score inferences in the contexts in which they are currently being used. The use of a variety of procedures to aid in the interpretation of test score results should be viewed as part of the ongoing process of construct validation.

Conclusions

Theoretically, this study has drawn upon L1 and L2 reading strategy research, cognitive psychology, SL assessment research, and psychometric research to develop a theoretical framework to test the hypothesis that some of the items included in the Canadian Language Benchmarks Reading Assessment favour certain cultural groups whose first language 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 and DBF analyses) to evaluate group differences on the CLBA Reading Assessment. By

employing a confirmatory approach to DIF in this study, valuable insight into the underlying causes of differential item and bundle functioning and the dimensionality of the CLBA Reading Assessment was gained.

Substantive analysis of the CLBA reading test items and DIF/DBF 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 favour 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 favour 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 a second language 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 modelling 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 may promote greater equity and fairness in CLBA reading comprehension test development practices. If the results of this study are confirmed in future research, test developers could use the information revealed about the bottom-up, top-down dimensions operating within the CLBA Reading Assessment to facilitate future item construction and the development of test specifications.

Multiple forms of evidence from additional confirmatory DIF, dimensionality, and multidimensional item response studies have the potential to illuminate the effects of linguistic/cultural background on the validity of CLBA 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.

References

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.

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.

Ackerman, T., Gierl, M., & Walker, C. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice, 22,* 37-53.

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, Louisiana.

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

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.

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,* Volume III (pp. 285-310). Mahwah, NJ: Erlbaum.

Allan, A. (1992). *EFL reading comprehension test validation: Investigating aspects of process approaches.* Unpublished PhD thesis, Lancaster University.

Anderson, J. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, N. (2002). The role of metacognition in second language teaching and learning. *ERIC Digest, April 2002,* 3-4.

Anderson, N. (1999). *Exploring second language reading: Issues and strategies.* Boston, MA: Heinle & Heinle.

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

Anderson, N., Bachman, L., Perkins, K., & Cohen, A. (1991). An exploratory study into the construct validity of a reading comprehension test: Triangulation of data resources. *Language Testing, 8,* 41-66.

Anderson, N., & Vandergrift, L. (1996). Increasing metacognitive awareness in the L2 classroom by using think-aloud protocols and other verbal report formats. In R. Oxford (Ed.). *Language learning strategies around the world: Cross-cultural perspectives* (pp. 3-18). University of Hawaii at Manoa: Second Language Teaching and Curriculum Center.

Angoff, W. & Ford, S. (1973). Item-race interaction on a test of scholastic aptitude.

Journal of Educational Measurement, 10, 95-106.

Argyris, C. (1970). *Intervention theory and mind: A behavioural science view.* Reading, MA: Addison-Wesley.

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

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

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

Bartlett, F. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, England: Cambridge University Press.

Bedell, D., & Oxford, R. (1996). 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.

Bernhardt, E. (1991). *Reading development in a second language.* Norwood, NJ: Ablex.

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.*

Block, E. (1992). See how they read: Comprehension monitoring of L1 and L2 readers. *TESOL Quarterly, 26,* 319-341.

Bloom, P., & Wynn, K. (1997). Linguistic cues in the acquisition of number words. *Journal of Child Language, 24,* 511-533.

Bolt, D. (2002). *Studying the DIF potential of nuisance dimensions using bundle DIF and*

multidimensional IRT analyses. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Bolt, D., & Stout, W. (1996). Differential item functioning: Its multidimensional model and resulting SIBTEST detection procedure. *Behaviormetrika, 23,* 67-95.

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, H. (1994). *Principles of language learning and teaching.* Upper Saddle River, NJ: Prentice Hall.

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

Brown, R. (1957). Linguistic determinism and the part of speech. *Journal of Abnormal and Social Psychology, 55,* 1-5.

Brown, R., & Pressley, M. (1994). Self-regulated reading and getting meaning from text: The transactional strategies instruction model and its ongoing validation. In D. Schunk & B. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications* (pp. 155–179). Hillsdale, NJ: Erlbaum.

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

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

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

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.

Chamot, A. (2004). Issues in language learning strategy research and teaching. *Electronic Journal of Foreign Language Teaching 1,* 14-26.

Chamot, A., & O'Malley, J. (1994). *The CALLA handbook: Implementing the Cognitive Academic Language Learning Instruction Approach.* White Plains, NY: Addison Wesley Longman.

Chamot, A., & Kupper, L. (1989). Learning strategies in foreign language instruction. *Foreign Language Annals, 22,* 13-24.

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

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.

Cohen, A., & Scott, K. (1996). Approaches to assessing language learning strategies. In R. Oxford (Ed.), *Language learning strategies around the world: Cross-cultural perspectives* (pp. 89-106). University of Hawaii at Manoa: Second Language Teaching and Curriculum Center.

Conrad, C. (1978). A grounded theory of academic change. *Sociology of Education, 51,* 101-112.

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

Douglas, J., Roussos, L., & Stout, W. (1996). Item-bundle DIF hypothesis testing: Identifying suspect bundles and assessing their differential functioning. *Journal of Educational Measurement, 33,* 465-484.

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., Vavrus, L.,

Wesselman, R., Putnam, J., & Bassiri, D. (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.

Ercikan, K., Gierl, M., McCreith, T., Puhan, G., & Koh, K. (2002). *Comparability of English and French Versions of SAIP for reading, mathematics and science items.* Paper presented at the annual meeting of the Canadian Society for the Study of Education, Toronto.

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.

Flavell, J. (1976). Metacognitive aspects of problem solving. In L. Resnick (Ed.), *The nature of intelligence* (pp. 231-235). Hillsdale, NJ: Erlbaum.

Gagné, E., Yekovich, C., & Yekovich, F. (1993). *The cognitive psychology of school learning* (2nd ed.). New York: HarperCollins.

Gierl, M., Bisanz, G., & Bisanz, J. (2001, July). *Developing an interpretative framework*

for understanding group differences on national and international achievement tests: The case of excellence in Alberta. A research proposal submitted to Alberta Learning, Edmonton, Alberta.

Gierl, M., Bisanz, J., Bisanz, G., Boughton, K., & Khaliq, S. (2001). Illustrating the utility of differential bundle functioning analyses to identify and interpret group differences on achievement tests. *Educational Measurement: Issues and Practice, 20,* 26-36.

Gierl, M., Bisanz, J., Bisanz, G., & Boughton, K. (2003). Identifying content and cognitive skills that produce gender differences in mathematics: A demonstration of the multidimensionality based DIF analysis framework. *Journal of Educational Measurement, 40,* 281-306.

Gierl, M., Rogers, T., & Klinger, D. (1999). *Consistency between statistical procedures and content reviews for identifying translation DIF.* Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Canada.

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. (1978). *Theoretical sensitivity.* Mill Valley, CA: Sociology Press.

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.

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

Grabe, W., & Stoller, F. (2002). *Teaching and researching reading.* London: Longman.

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

Gutierrez, K., & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher, 32*(5), 19-25.

Hacker, D. (1998). Self-regulated comprehension during normal reading. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 165–191). Mahwah, NJ: Erlbaum.

Hammadou, J. (1991). Interrelationships among prior knowledge, inference, and language proficiency in foreign language reading. *Modern Language Journal, 75,* 27-38.

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.

Hart, E., & Speece, D. (1998). Reciprocal teaching goes to college: Effects for postsecondary students at risk for academic failure. *Journal of Educational Psychology, 90,* 670–681.

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.

Holland P., & Thayer, D. (1986). *Differential item performance and the Mantel-Haenszel procedure.* ETS Research Report No. 86-31. Princeton, NJ: Educational Testing Service.

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

Howard, D. (1985). *Cognitive psychology.* New York: Macmillan.

Jiang, H., & Stout, W. (1998). Improved Type I error control and reduced estimation bias for DIF detection using SIBTEST. *Journal of Educational and Behavioural Statistics, 23,* 291-322.

Jones, W. (1980). Newcomers' biographical explanations: The self as an adjustment process. *Symbolic Interaction, 3,* 83-94.

Kagitçibasi, C. (1996). *Family and human development across cultures: A view from the other side.* Hillsdale, NJ: Erlbaum.

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.

Kim, M. (2001). Detecting DIF across the different language groups in a speaking test. *Language Testing, 18,* 89-114.

Kinsche, W. (1986). Learning from text. *Cognition and Instruction, 3,* 87-108.

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.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessment. *Journal of Personality and Social Psychology, 77,* 1121-1134.

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). 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)

Magliano, J., Trabasso, T., & Graesser, A. (1999). Strategic processing during comprehension. *Journal of Educational Psychology, 91*, 615–629.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22,* 719-748.

Madaus, G. (1990). *Testing as social technology. The inaugural annual Boise lecture on*

education and public policy. Boston, MA: Boston College.

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

Meyer, B. (1975). *The organization of prose and its effects on memory.* New York: North Holland.

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

Nandakumar, R. (1993). Simultaneous DIF amplification and cancellation: Shealy-Stout's test for DIF. *Journal of Educational Measurement, 30,* 293-311.

Nayak, N., Hansen, N., Krueger, N., & McLaughlin, B. (1990). Language-learning strategies in monolingual and multilingual adults. *Language Learning, 40,* 221-244.

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

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

O'Malley, J., Chamot, A., Stewner-Manzanares, G., Russo, G., & Kupper, L. (1985). Learning strategy applications with students of English as a second language. *TESOL Quarterly, 19,* 285-296.

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

Palincsar, A., & Brown, A. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1,* 117–175.

Paris, S., Lipson, M., & Wixson, K. (1983). Becoming a strategic reader. *Contemporary*

Educational Psychology, 8, 293-316.

Paris, S., Wasik, B., & Turner, J. (1991). The development of strategic readers. In R. Barr, M. Kamil, P. Mosenthal, & P. Pearson (Eds.), *Handbook of reading research:* Vol. 2. (pp. 609-640). New York: Longman.

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.

Pressley, M., El-Dinary, P., & Brown, R. (1992). Skilled and not-so-skilled reading: Good information processing and not-so-good information processing. In M. Pressley, K. Harris, & J. Guthrie (Eds.), *Promoting academic competence and literacy in school* (pp. 91–127). San Diego, CA: Academic Press.

Principles for Fair Student Assessment Practices for Education in Canada. (1993). Joint Advisory Committee. Retrieved October 9, 2003, from University of Alberta, Centre for Research in Applied Measurement and Evaluation Web site: http://www.education.ualberta.ca/educ/psy/crame/research.htm

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.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading.* Englewood Cliffs, NJ: Prentice Hall.

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.

Rogers, W. T., & Yang, P. (1996). Test-wiseness: Its nature and application. *European Journal of Psychological Assessment, 12,* 247-259.

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

Routman, R. (1994). *Invitations: Changing as teachers and learners K-12.* Portsmouth, NH: Heinemann.

Rumelhart, D. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance.* New York: Academic Press.

Rumelhart, D. (1980). Schemata: The building blocks of cognition. In R. Shapiro, B. Bruce & W. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 123-156). Hillsdale, NJ: Erlbaum.

Ryan, A., & Meara, P. (1991). The case of the invisible vowels: Arabic speakers reading English words. *Reading in a Foreign Language, 7,* 531-40.

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

Sarig, G. (1987). High-level reading in the first and in the foreign language: Some comparative process data. In J. Devine, P. Carrell, & D. Eskey (Eds.), *Research in reading in English as a second language* (pp. 105-120). Washington: TESOL.

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

Schank, R., & Abelson, R. (1977). *Scripts, plans, goals and understanding.* Hillsdale, NJ: Erlbaum.

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

Schueller, J. (2000). The effects of two types of strategic training on foreign language reading comprehension. An analysis by gender and proficiency. *Dissertation Abstracts International, 60* (07), 2472. (UMI No. 9923247)

Schueller, J. (in press). Gender and foreign language reading comprehension: The effects of strategy training. *Southern Journal of Linguistics.*

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.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

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

Stanovich, K. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16*, 32-71.

Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-407.

Stanovich, K. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. New York: Guilford Press.

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

Storti, C. (1989). *The art of crossing cultures.* Yarmouth, MA: Intercultural Press.

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

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

Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of

group differences in trace lines. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 147-169). Hillsdale, NJ: Erlbaum.

Thompson-Panos, K., & Thomas-Ružić, M. (1983). The least you should know about Arabic: Implications for the ESL writing instructor. *TESOL Quarterly, 17,* 609-623.

van Dijk, T. (1977). *Text and context: Explorations in the semantics of text.* London: Longman.

Vann, R., & Abraham, R. (1990). Strategies of unsuccessful language learners. *TESOL Quarterly, 24,* 177-194.

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.

Widdowson, H. (1983). *Learning purpose and language use.* Oxford: Oxford University Press.

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

Wolff, D. (1987). Some assumptions about second language text comprehension. *Studies in Second Language Acquisition, 9,* 307-326.

Yang, P. (2000). Effects of test-wiseness upon performance on the Test of English as a Foreign Language (Doctoral dissertation, University of Alberta, 2000). *Dissertation Abstracts International, 62,* 1724.

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.

Zumbo, B. (1999). *A handbook on the theory and methods of differential item functioning: Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores.* Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Appendix A

Invitation to Participate in an ESL Reading Study

Dear Student:

I am writing this letter to introduce myself to you and to invite you to participate in a study that I am conducting for my dissertation. I am a PhD student who is studying Applied Measurement and Evaluation in the Department of Educational Psychology at the University of Alberta.

I am only recruiting participants who are intermediate ESL students and whose first language is either Arabic or Mandarin.

By having the opportunity to interview you and having you answer some reading comprehension questions, I hope to understand more about how people read in a second language. My desire to explore this topic stems from my personal and professional interests in second language learning and testing. By undertaking this research, I hope to increase my understanding of the processes involved in second language reading.

Your participation in this study will be in the form of two reading assessment sessions with myself and another person who speaks your native language. Both sessions will take place on a weekday at the college. During this first session, I will ask you to fill out a background information questionnaire. Then I will ask you to answer 18 reading comprehension questions. While you are answering the questions, I would like you to tell me what you are thinking and what you are focusing on in the text that helps you to answer each question. I will also ask about your understanding of each question and each reading passage. Your responses will be audio-taped. During the second session, using the same procedures, I will ask you to complete 18 additional reading comprehension questions. Then I will ask you to answer some questions about what you do when you read. This study will take approximately three hours of your time.

There will be no risks to you as a participant, and you will be free to withdraw from the study at any time for any reason. Withdrawing from the study will not have any impact on your status as a student at the college. You will remain anonymous – no names will be collected on the data sheets. I will keep the data secure for five years after completion of the study – at that point, the computer and audio files will be destroyed and the test booklets, transcripts, and questionnaires will be shredded. Only the group results will be discussed in the research paper that I will write, present at conferences, or use for the purposes of teaching.

If you have any questions or if you would like to discuss anything with me, or my supervisor, please feel free to contact either of us via email or at the phone numbers below. If you would like to participate in the study please notify your instructor. Then I will contact you to set up two afternoon appointments.

Yours truly,

Marilyn Abbott
PhD Candidate
Department of Educational Psychology
University of Alberta

Dr. Tracey Derwing
Professor
Department of Educational Psychology
University of Alberta
Tel. 492-3668
tracey.derwing@ualberta.ca

This study has been reviewed and approved by the Faculties of Education and Extension Research Ethics Board (EE REB) at the University of Alberta. For questions regarding participant rights and ethical conduct of research, contact the Chair of the EE REB at (780) 492-3751.

Appendix B

Language Background Questionnaire

Date: _____ #_____

1. What is your native language (mother tongue)? _____

2. How old are you? _____ years

3. ❑ Male ❑ Female

4. What is your level / years of education? school ____ technical ____ university/college ____

5. What was your occupation in your first country? _____

6. Where were you born? (city, country) _____

7. What is your mother's first language? _____

8. What is your father's first language? _____

9. How long did you study English in your first country? (years and months) _____

10. How long have you been studying English in Canada? (years and months) _____

11. How much time have you spent in an English speaking country? _____

11. List any languages other than your native language that you read fluently.

_____

12. List any languages you read a little. _____

Appendix C

Participant # _____

READING ABILITY, PREFERENCES, AND STRATEGIES QUESTIONNAIRE: Indicate the number of the response that best describes you:

1 = strongly disagree, 2 = disagree, 3 = unsure, 4 = agree, 5 = strongly agree

1. You are a good reader in English.

   1 2 3 4 5

2. When answering the CLBA reading comprehension questions:

   a. you were able to recognize the difference between the main points and supporting details.

      1 2 3 4 5

   b. you were able to relate information which comes next in the text to previous information in the text.

      1 2 3 4 5

   c. you focused on understanding the meaning of each word.

      1 2 3 4 5

   d. you focused on the grammatical structures.

      1 2 3 4 5

   e. you focused on relating the text to what you already knew about the topic.

      1 2 3 4 5

   f. you focused on the details of the content.

      1 2 3 4 5

   g. you focused on the organization of the text.

      1 2 3 4 5

   h. you broke words into smaller parts to help you understand their meaning.

      1 2 3 4 5

   i. you tried to find synonyms for words in the text.

      1 2 3 4 5

3. When reading silently in English:

   a. if you don't understand something, you look up unknown words in a dictionary.

   1 2 3 4 5

   b. you focus on mentally sounding out parts of the words.

   1 2 3 4 5

   c. you focus on being able to pronounce each word silently to yourself.

   1 2 3 4 5

4. You are a good reader in your native language.

   1 2 3 4 5

5. You like to learn English by studying grammar.

   1 2 3 4 5

6. You like to learn English by talking to native speakers.

   1 2 3 4 5

7. You like to learn English by reading magazines and books.

   1 2 3 4 5

8. You like to learn English by writing; for example, writing essays, stories, using email.

   1 2 3 4 5

Thank you for participating in this study!

Appendix D: Confidentiality Agreement

This form was used for individuals hired to assist with interviews, and where necessary, to translate and transcribe the verbal report data conducted at the college.

Project title: English Reading Strategies: Differences in Arabic and Mandarin Speaker Performance on the CLBA Reading Assessment

I, _____, a bilingual interpreter have been hired to assist Marilyn Abbott with her data collection procedures. I understand that I will be required to help conduct interviews, and translate and transcribe verbal reports.

I agree to:

1. keep all the research information shared with me confidential by not discussing or sharing the research information in any form or format (e.g., computer disks, tapes, transcripts, CLBA items) with anyone other than Marilyn Abbott.

2. keep all research information in any form or format (e.g., disks, tapes, transcripts) secure while it is in my possession.

3. return all research information in any form or format (e.g., disks, tapes, transcripts) to Marilyn Abbott when I have completed the research tasks.

4. after consulting with Marilyn Abbott erase or destroy all research information in any form or format regarding this research project that is not returnable to Marilyn Abbott (e.g., information stored on a computer hard drive).

_____	_____	_____
(print name)	(signature)	(date)
Marilyn Abbott	_____	_____
(print name)	(signature)	(date)

This study has been reviewed and approved by the Faculties of Education and Extension Research Ethics Board (EE REB) at the University of Alberta. For questions regarding participant rights and ethical conduct of research, contact the Chair of the EE REB at (780) 492-3751.

Appendix E

ESL Participant Consent Form

This study is being conducted to help me understand more about how people read in a second language. Your participation is greatly appreciated. There are no risks to you as a participant, and you are free to withdraw from the study at any time for any reason. Withdrawing from the study will not have any impact on your status as a student. You will remain anonymous - - no names will be collected on the data sheets. Only the group results will be discussed in the paper that I will write. I will need to interview you on two separate occasions. On day one, I will ask you to fill out a background information questionnaire. Then I will ask you to answer 18 reading comprehension questions. While you are answering the questions, I would like you to tell me what you are thinking and what you are focusing on in the question or passage that helps you to get the answer. I will also ask about your understanding of the questions and the passages. Your responses will be audio-taped. On day two, using the same procedures, I will ask you to answer 18 reading comprehension questions. After you have completed the questions, I will ask you to answer 18 questions about what you do when you read. This study will take approximately three hours of your time.

Please circle yes or no for each of the following statements:

- I consent to voluntary participation in this study. Yes / No
- I understand that I can withdraw at any time. Yes / No
- I understand that there are no risks involved. Yes / No
- I understand that I will be ask to provide some information on questionnaires and answer some reading comprehension questions. Yes / No
- I understand that my responses will be audio-recorded. Yes / No

Name: _____
(please print)

Signature: _____

Date: _____

Marilyn Abbott
PhD Provisional Candidate
Department of Educational Psychology
University of Alberta

Dr. Tracey Derwing
Professor
Department of Educational Psychology
University of Alberta
Tel. 492-3668
tracey.derwing@ualberta.ca

This study has been reviewed and approved by the Faculties of Education and Extension Research Ethics Board (EE REB) at the University of Alberta. For questions regarding participant rights and ethical conduct of research, contact the Chair of the EE REB at (780) 492-3751.

Appendix F

Concurrent Verbal Report Practice Instructions Adapted from Green (1998)
Note: Verbal Instructions are in Bold

In this study we are interested in what you think about while you are reading and answering some questions about a passage. To find out about this, we are going to ask you to think aloud as you read and work through the questions. By think aloud we mean that we want you to say out loud everything that you are thinking in whatever language you are thinking in from the time you start reading the question until you select an answer. We would like you to talk constantly from the time you start reading and answering the first question until you select your answer. It is important that you do not plan out or try to explain to us what you are thinking. It may help to imagine that you are in the room by yourself. It is very important that you keep talking. If you are silent for any period of time, we will remind you to keep talking. Do you understand what we are asking you to do?
Let's begin with a few practice questions (CLBA Form 2, Stage II, Task A).

- Ensure that the participants do not provide activity descriptions as activity descriptions such as "I'm just reading this paragraph here" are not the same as thinking or talking aloud (Green, 1998).
- Also ensure that they do not merely focus on reporting their test-wiseness strategies.
- No model was given so as to avoid bias by participants imitating the model given and to ensure spontaneity of the response (Green, 1998).

Once they have finished the first question provide them with instructions on how to report retrospectively. Ericsson and Simon (1993) recommend that individuals practice reporting both ways to emphasize the difference between concurrent and retrospective reporting and so the investigator can compare the contents of both reports as both should contain roughly the same information. In essence, the reading strategies that are reported introspectively will be validated retrospectively.

Retrospective Verbal Report Instructions Adapted from Green (1998)

Now we would like you to tell us what you can remember about what you were thinking and what you were attending to from the time you read the practice question until you gave us your answer. We are interested in what you can actually remember, not what you think you may or should have thought. If possible, it would be best if you can tell us what you remember in the order in which you memories occurred as you worked through the question. If you are not sure about any of your memories, please say so. We do not want you to try to answer the question again, we just want you to tell us what you can remember thinking and what you were attending to when you were reading and answering the question. Now tell us what you can remember.

- While they are thinking aloud write down any descriptions regarding their non-verbal behaviour.

Verbal Instructions for the Form 1, Stage II CLBA Passages

1. **Now you will be given a passage with corresponding questions to answer.**
2. **While reading and answering each question you will be asked to think aloud as you did in the practice session. We would like you to say out loud everything that you are thinking and focusing on from the time you start reading the question until you select an answer.**
3. **You will not be interrupted or assisted once you begin.**
4. **If you pause for any length of time, you will be reminded to keep talking.**
5. **Once you have answered the question, we want you to tell us all that you can remember about your thinking and what you were attending to when reading and answering the question.**

- Remember to record their non-verbal behaviours.

- After they have completed the questions for the passage have them fill in the passage and item comprehension rating form.

Appendix G: Confidentiality Agreement

This form was used as a confidentiality agreement between Marilyn Abbott and the immigrant referral centres.

Project title: English Reading Strategies: Differences in Arabic and Mandarin Speaker Performance on the CLBA Reading Assessment

I, Marilyn Abbott agree to:

1. keep all the research information shared with me confidential by not discussing or sharing the research information in any form or format (e.g., client information, CLBA reading items, computer data).

2. keep all research information in any form or format (e.g., client information, CLBA test items, computer data) secure while it is in my possession.

3. only enter sample identification numbers into the computer. Neither the clients' names nor their file numbers will be recorded.

4. return all research information in any form or format (e.g., client information, CLBA test forms) to their appropriate files once I have completed entering the demographic and item level data.

5. securely store the data. According to SSHRC guidelines, I will keep the data secure for five years after completion of the study – at that point, all data will be destroyed.

Marilyn Abbott	_____	_____
(print name)	(signature)	(date)

This study has been reviewed and approved by the Faculties of Education and Extension Research Ethics Board (EE REB) at the University of Alberta. For questions regarding participant rights and ethical conduct of research, contact the Chair of the EE REB at (780) 492-3751.

Appendix H: Confidentiality Agreement

This form was used for the individual hired to enter demographic and item level CLBA test data at the immigrant referral centres.

Project title: English Reading Strategies: Differences in Arabic and Mandarin Speaker Performance on the CLBA Reading Assessment

I, _____, have been hired to assist Marilyn Abbott with her data collection procedures. I understand that I will be required to help enter demographic and item level data from already existing Canadian Language Benchmark Assessment forms.

I agree to:

1. keep all the research information shared with me confidential by not discussing or sharing the research information in any form or format (e.g., client information, CLBA test items, computer data) with anyone other than Marilyn Abbott.

2. keep all research information in any form or format (e.g., client information, CLBA test items, computer data) secure while it is in my possession.

3. return all research information in any form or format (e.g., client information, CLBA test items, computer data) to Marilyn Abbott when I have completed the research tasks.

4. after consulting with Marilyn Abbott erase or destroy all research information in any form or format regarding this research project that is not returnable to Marilyn Abbott (e.g., information stored on a computer hard drive).

_____	_____	_____
(print name)	(signature)	(date)
Marilyn Abbott	_____	_____
(print name)	(signature)	(date)

This study has been reviewed and approved by the Faculties of Education and Extension Research Ethics Board (EE REB) at the University of Alberta. For questions regarding participant rights and ethical conduct of research, contact the Chair of the EE REB at (780) 492-3751.

Appendix I

ESL Expert Consent Form

The purpose of this study is to understand how reading strategies interact with linguistic/cultural background to affect test performance on the Canadian Language Benchmarks Assessment (CLBA) reading subtest. Your participation is greatly appreciated. There are no risks to you as a participant, and you are free to withdraw from the study at any time for any reason. You will remain anonymous - - no names will be collected on the data sheets. Only the group results will be discussed in the paper that I will write. After a training session to introduce you to the reading strategies coding schema, you will be asked to classify 32 CLBA questions into seven bottom-up and five top-down reading strategy categories. Thus you will code the items according to the reading strategy that you believe will be the most instrumental in arriving at the correct answer and then rate the usefulness of each strategy for answering each of the CLBA reading items. Following the independent coding session, a meeting will be held so all three expert judges can reach a consensus regarding the coding of the items on which they disagree. This study will take approximately six hours of your time. The consensus codings developed in this part of the study will be used to group the items and conduct differential item and bundle functioning analysis of group differences in reading strategies on the CLBA.

Name: _____
(please print)

Signature: _____

Date: _____

Marilyn Abbott
PhD Provisional Candidate
Department of Educational Psychology
University of Alberta

Dr. Tracey Derwing
Professor
Department of Educational Psychology
University of Alberta
Tel. 492-3668
tracey.derwing@ualberta.ca

This study has been reviewed and approved by the Faculties of Education and Extension Research Ethics Board (EE REB) at the University of Alberta. For questions regarding participant rights and ethical conduct of research, contact the Chair of the EE REB at (780) 492-3751.

Appendix J

CLBA READING ASSESSMENT CODING SHEET

INSTRUCTIONS: Please use the following coding sheet to summarize your ratings of each CLBA reading comprehension item. Items should be coded using the reading strategy classification schema outlined below. Be sure to have a copy of the item in front of you to assist you when identifying the most applicable strategies. First, indicate WHICH reading STRATEGY OR STRATEGIES would be most instrumental in answering each item. Second, evaluate the SALIENCY (i.e., importance) of each strategy. Third, identify which strategy is MOST SALIENT. Finally, indicate whether each item is a GOOD/CLEAR EXAMPLE of an item that elicits the most salient strategy you identified.

A. ITEM NUMBER: Task ____ Item ____

B. ITEM CHARACTERISTICS:

	WHICH? (Choose all that apply)	SALIENCY[1] 1 = Not at all salient 2 = Not very salient 3 = Salient 4 = Very salient	MOST SALIENT	GOOD EXAMPLE (of most salient)
B1. breaks lexical items into parts				
B2. scans for explicit information requested in the item				
B3. identifies a synonym or a paraphrase of the literal meaning of a word, phrase, or sentence				
B4. relates verbal information to accompanying visuals				
B5. matches key vocabulary in the item to key vocabulary in the text				
B6. uses knowledge of grammar or punctuation				
B7. uses local context cues to interpret a word or phrase				
T1. skims for gist/identifies the main idea, theme, or concept				
T2. connects or relates information presented in different sentences or parts of the text				
T3. draws an inference based on information presented in the text				
T4. uses background knowledge to speculate beyond the text				
T5. recognizes discourse format				

[1]Please answer the following question: How salient is this strategy either in the item or in how likely examinees are to use it to answer the question?