EXPLORING FACTORS ASSOCIATED WITH ESL/ELD STUDENTS'

PERFORMANCE ON THE ONTARIO SECONDARY SCHOOL LITERACY TEST

by

YING ZHENG

A thesis submitted to the Faculty of Education

In conformity with the requirements for

the degree of Master of Education

Queen's University

Kingston, Ontario, Canada

December, 2005

Abstract

The study explored factors associated with English as Second Language (ESL) and English Literacy Development (ELD) students' performance on the Ontario Secondary School Literacy Test (OSSLT) from three different yet interrelated aspects. These aspects were the relationship between test performance and the characteristics of ESL/ELD students; the relationship between ESL/ELD students' aggregated school performance and their school-level socio-economic status; and a comparison of ESL/ELD and non-ESL/ELD students' performance on three test formats in the reading component of the OSSLT (multiple-choice questions, constructed-response questions, and constructed-response questions with explanations). The study was conducted based on 4,311 ESL/ELD students' test data, their responses to the Educational Quality and Accountability Office's (EQAO) Student Questionnaire, and the test results of 5,003 non-ESL/ELD students. School-level socio-economic status data from the Education Quality and Indicators Program (EQUIP) were merged with the OSSLT test data. These data were analyzed using exploratory factor analysis, multiple regression and discriminant analysis.

The results indicated that e-literacy activities, literature literacy activities, non-fiction literacy activities, newspaper and magazine literacy activities, and literacy hours were positive predictors of ESL/ELD students' performance on the OSSLT. Letter reading and writing, song and poetry literacy activities were two negative predictors. In terms of home language patterns, whether English was the first language of the test-takers, and what language(s) they spoke at home were both influential variables in differentiating ESL and ELD students' literacy performance. The results also revealed that a higher parental education index positively predicted a school's average OSSLT performance. In

addition, ESL/ELD students demonstrated substantial performance gaps in all of the three

test formats in reading as compared to non-ESL/ELD students. Only multiple-choice

questions obtained a significant discriminant coefficient with a weak discriminating

function. The results of this study offered some insights about identifying and

understanding factors that were associated with ESL/ELD students' OSSLT performance

from the perspective of the test-takers and the test itself. The results also provide

directions for future research and instructional support in relation to ESL/ ELD students

in the context of the accountability framework in Ontario.

Acknowledgements

The first people that should be acknowledged for their roles in this thesis are the members of my thesis committee. I owe an immense debt of gratitude to Dr. Liying Cheng, my supervisor, for her knowledgeable and helpful suggestions and for her constant encouragement and support. I am also grateful for Dr. Don A. Klinger, who offers me insightful guidance in the whole thesis writing process. Both of them read and re-read drafts of this study and offered extensive and thoughtful comments. These comments and the patience that they displayed were very much appreciated. I would also like to thank Dr. Nancy Hutchison, the coordinator of the program, who is always there and ready to offer timely support and generous help whenever I need it.

My family has been especially supportive of my efforts to pursue this study, exhibiting their extraordinary patience and understanding. I also want to thank Marlene, Heather, and the professionals in the Education library. My gratitude also goes to my friends and colleagues who have shared all the joys and tears with me in the past two years.

Table of Contents

List of Tables

List of Figures

Chapter One: Introduction

The past two decades have witnessed a dramatic increase in the numbers of immigrants accepted by Canada. Canadian Immigration and Citizenship (2003) showed that in 1985, the numbers of immigrants were 84,335; forming 0.3% of the population. In 2002, the numbers quickly reached 229,091, forming 0.7% of the population. Immigrants come with their families, therefore, the numbers of school-aged immigrant students have expanded simultaneously across the nation (reached 82,569 in 2002) (Canada Immigration and Citizenship, 2002). Students speaking English as second language (ESL) present ethnic, linguistic, and cultural diversities (Watt & Roessingh, 2001). In Ontario high schools, this phenomenon is represented by the increasing proportions of ESL and ELD (English Literacy Development) students[1]. Approximately one in four students who wrote the October 2003 Ontario Secondary School Literacy Test (OSSLT) reported that their first language was a language other than English (EQAO, 2004). Presently, as evidenced by the low success and high deferral rates on the OSSLT, many of the ESL/ELD students are encountering great difficulties as they attempt to obtain fundamental English literacy skills.

The OSSLT is a province-mandated standardized test in English literacy, which is a graduation requirement for all Ontario secondary school students in order to receive their secondary school diploma. Statistics show a discernible gap in the success rates of ESL/ELD students and all other students (students can have the option of deferring the

---

[1] ESL students are students whose first language is not English, but have received educational experiences in their own countries using their first language. ELD students are those who are from countries where access to education may have been limited, and they have few opportunities to develop literacy skills in any language (Ministry of Education and Training, 1999, p. 2). Although they are treated as one group of students except in the analysis of their performance on the test formats compared with non-ESL/ELD students, they are two groups of students with distinct characteristics in their learning and testing performance.

test). For example, in the previous three test administrations, only 37% in February 2002, 34% in October 2002, and 42% in October 2003 of the ESL/ELD students passed the whole test. In contrast, for all the students who wrote the test, the overall pass rate was 75%, 72% and 77% in the above-mentioned three test administrations. Research has demonstrated that it is of pedagogical significance to focus on ESL/ELD students who are still in their language "developmental process" to gain academic achievement (Collier, 1989). Research has also indicated that test-takers with different characteristics might be affected by a test in ways that are not relevant to the abilities being tested (Bachman, 2000; Kunnan, 1998). This study, therefore, attempted to identify factors that were associated with ESL/ELD students' performance on the OSSLT. The factors focused on both the test-takers and the test itself. The examination of the test-takers included test-takers' individual characteristics and their average school socio-economic (SES) background. The examination of the test mainly focused on test format differences in the reading section of the OSSLT.

Research Context

The study was designed to examine ESL/ELD students who wrote the OSSLT in October 2003. The OSSLT, which the Ministry of Education introduced in 1999, is administered by the Education Quality and Accountability Office (EQAO), the arm's-length agency of the provincial government. The OSSLT is a minimal competency test, which means literacy on the OSSLT comprises the reading and writing skills that are required to understand reading selections and to communicate through a variety of written forms as expected in the Ontario Curriculum across all subjects up to the end of grade 9. There have been four test administrations so far: February 2002, October 2002, October

2003 and October 2004. To meet the requirements of the Ontario Secondary School Diploma (OSSD), all students must successfully complete both the reading component and the writing component of the test. Students who have been previously eligible to write the test at least twice but have not yet been successful at least once are eligible to fulfill the requirement through the Ontario Secondary School Literacy Course (OSSLC). For these students, successful completion of the OSSLC can fulfill the graduation requirement associated with the OSSLT (EQAO, 2002).

As mentioned earlier, there are increasing numbers of ESL/ELD students in Ontario high schools, who bring with them increasing linguistic and cultural diversity. It is necessary to know clearly how ESL students and ELD students are defined in Ontario. ESL students are those who have recently arrived in Ontario schools "from a country with a first language other than English and with little or no previous experience with English". They are the students who may be proficient users of their own language(s). ELD students are "those who arrive in Canada from countries where English is the first language but other varieties of English are in common use" (EQAO, 2002, p. 7). Two distinctions are acknowledged between ESL students and ELD students within this context. The first is whether English is the student's first language. The second is whether the student has received formal schooling before entering the Ontario system. More specifically, ESL students may have received educational experiences in their own countries, which means that they have had a chance to develop their literacy skills in their first language. ELD students are usually from countries or communities where the access to education has been limited, or they may have had few or no opportunities to develop literacy skills in any language, but have sufficient knowledge of oral English to benefit from literacy instruction in English (Ministry of Education and Training, 1999).

In order to make sure that the actual OSSLT test items are grade- and age-appropriate, EQAO states that the actual items are reviewed by committees of educators (EQAO, 2002). The test is regarded as a useful quality assurance, which shows "the extent to which Ontario students are meeting a common, basic standard for literacy across the province" (EQAO, 2002, p.1). The standards created for the OSSLT are descriptions of student performance that constitute a minimum acceptable level of achievement in English language literacy when a student reaches grade 10. Descriptions of student work in reading and writing are developed to demonstrate performance that meets or exceeds the minimum criteria (passed), or that does not meet the criteria (did not pass). ESL/ELD students are evaluated using the same standards as their non-ESL/ELD counterparts. Many ESL/ELD students may have only been in the Ontario system for a limited period of time before they write the OSSLT. Researchers in second language education suggest that four to eight years is required for ESL/ELD students to attain a level of language proficiency necessary to compete on a minimally competent level with their non-ESL/ELD counterparts (Collier, 1989; Cummins, 1981; Roessingh, 1999). However, if ESL/ELD students are not allowed the time to make up for their English language development, the consequence is that they will be more likely to fail the test, which may potentially have negative consequences on their academic studies.

Coupled with the high failure rates, ESL/ELD students also chose to defer from writing the test at far higher rates than the rest of the population. For example, of all the ESL/ELD students who were eligible to write the test, only 46% in February 2002, 45% in October 2002, and 54% in October 2003 participated in the test. Meanwhile, the participation rates of all the students were 91%, 93% and 91% respectively (EQAO, 2003). This study is significant in understanding ESL/ELD students' performance on the

OSSLT and exploring what factors may contribute to or impede their success on the test. Before addressing the research questions and the rationale for the study, the OSSLT components and their scoring criteria are presented in the next section for the purpose of providing more relevant information regarding the research context.

OSSLT Components and Scoring Criteria

The OSSLT has two components: reading and writing. In the reading component, there are 100 questions about 12 different reading selections based on three types of texts: information (50%), consisting of explanation and opinion; graphic (25%), consisting of graphs, schedules and instructions; narrative (25%), consisting of stories and dialogues. The students are expected to demonstrate the following three reading skills as required: understanding directly stated ideas and information; understanding indirectly stated ideas and information; making connections between personal experiences and information in a reading selection. Finally, there are three kinds of test formats in the reading component with a total of 200 marks, namely, multiple-choice questions (40 questions, 80 marks), constructed-response questions (35 questions, 70 marks), and constructed-response questions with explanations (25 questions, 50 marks). The cut score for reading is 125. In the writing component, four types of writing tasks are included: a summary, a series of paragraphs expressing an opinion, a news report and an information paragraph. The writing score had a total of 180 marks with 100 as the cut score. As mentioned earlier, the standards created for the OSSLT are descriptions of the student performance that constitute a minimum acceptable level of achievement. The student who reads with "reasonable accuracy and proficiency" can pass the reading section. In other words, he or she "usually connects relevant ideas and information to understand the meaning" and "has

moderate success in integrating personal knowledge and experience to extend meaning."
In terms of the criteria to pass the writing component, he or she needs to demonstrate "an ability to develop, organize and connect main ideas with some supporting information;" "generally uses appropriate words and sentence structure;" and "generally does not make errors in spelling, grammar and punctuation that impede meaning" (EQAO, 2002, p.7-8).

With respect to the scoring criteria in the reading component, all questions are marked according to a well-defined scoring instrument. Multiple-choice (MC) questions are scored separately. Written answers are scored using an Answer Key for each question. Constructed-response (CR) questions ask students to respond in a few words to correctly answer the question. Two points were given for each item answered correctly, and no partial credit is given. Incorrect answers for each item are given 0 points. Constructed-response with explanation (CRE) questions ask students to justify or explain the thinking behind their answers. Answers are marked correct (2 points), partly correct (1 point), or incorrect (0 points) (EQAO, 2003). In addition, writing is marked using a holistic scoring method, which looks at the overall quality of the writing by examining the degree to which the students integrate the elements of writing in meeting the specific demands of the writing tasks. Writing quality is scored from a low of 0 to a high of 4 points (0-4). All scores in each writing task are weighted accordingly. The weighting scheme is as follows: 0 is converted to 0, 1 point is converted to 10 points, 2 points is converted to 20 points, 3 points is converted to 35 points, and 4 points is converted to 45 points. The total writing weighted score (total writing score thereafter) is the weighted mark earned by a student by adding the four weighted writing scores for the four writing tasks.

Research Questions

Previous research has indicated that students' test performance may not be a pure demonstration of their abilities (Bachman, 1990; Kunnan, 1995). It may, in most cases, interact with factors that are derived from test-takers, e.g., personal attributes and family backgrounds. Also, test performance may interact with factors that are derived from the test itself, e.g., test format differences. This study attempted to examine the relationship between ESL/ELD students' OSSLT performance and its interacting factors, including test-taker characteristics, average school-level socio-economic status, and test format differences. The following three research questions were examined:

1) What relationships, if any, exist between ESL/ELD students' performance on the OSSLT and their individual test-taker characteristics? How do test-taker characteristics predict individual performance?

2) What relationships, if any, exist between ESL/ELD students' average school performance on the OSSLT and the average school-level socio-economic indicators? How do socio-economic indicators predict ESL/ELD students' school performance?

3) What are the performance differences, if any, between ESL/ELD students and non-ESL/ELD students in the reading component of the OSSLT? How do test formats discriminate between ESL/ELD students and non-ESL/ELD students?

Rationale

Given the fact that ESL/ELD students displayed substantial performance gaps on the OSSLT, this study aimed to gain an in-depth understanding of the factors that were

associated with these students' performance. The factors informed two issues: test-takers characteristics and test formats. The rationale for conducting this research is presented in the following section.

Research has suggested that powerful as testing is, it has brought about both intended and unintended consequences to different groups of students (Madaus & Clarke, 2001). Minority students such as ESL/ELD students are among the groups most vulnerable to the effects of high-stakes testing policies (Shepard, 1991). High-stakes tests tend to have more severe consequences for minority students and students from poor families (Horn, 2003; Madaus & Clarke, 2001), although minority students in these research contexts include black and Hispanic students, as well as ESL students. It should also be noted that in terms of English literacy development, ESL/ELD students may not be in an equal position to their non-ESL/ELD counterparts, who have likely been part of the Canadian educational system for most if not all of their education (Diao, 2004). Another reason why ESL/ELD students may be adversely affected by tests is that these tests are originally designed for their non-ESL/ELD counterparts. Cornell (1995) has also argued that evaluation criteria that heavily rely on English language skills overlook language minority students' individual progress, and are apt to fail them in tests.

Moreover, the ubiquitous debate in education involves the identification of factors that contribute to student achievement (Mulvenon, Ganley, & Fritts-Scott, 2001). According to Messick (1989), in order to achieve test validity, it is fundamental to make sure that low test scores are not a result of factors that are irrelevant to the constructs an instrument intends to measure (Messick, 1989). Bachman (1990) has proposed three broad categories that could affect performance on language tests. They include test method facets, attributes of the test-takers and random factors. He argues that in language

testing, in order for a test score to be a meaningful indicator of a particular individual's ability, we must be sure it measures students' language abilities and very little else (Bachman, 1990). However, test performance may also be affected by individual attributes that are not part of a test-taker's language ability and by characteristics of the methods used to elicit test performance. Test-taker characteristics (personal attributes) consist of "cultural background, background knowledge, cognitive abilities, sex, and age" (Bachman 1990, p. 350). The test method facet is a category which briefly includes the testing environment, the test rubrics, the nature of the input the test-takers received, the nature of the expected response to that input, and the relationship between input and response (Bachman, 1990, p. 119). The third category that influences language test performance is random measurement error, which is made up of interactions between components of students' abilities, test method facets, personal characteristics, and measurement error.

Bachman (1990) has reiterated that a major concern in the design and development of language tests is to minimize the effects of test methods, personal attributes that are not part of students' abilities, and random factors on test performance. In addition, the investigation of test method facets are of particular importance for designing, developing, and using language items because contrary to the test-taker characteristics, which are beyond the control of teachers or test designers, the test method facet is something that is potentially controllable in real testing practices (Bachman, 1990).

In addition to Bachman's theoretical discussion on the factors interacting with students' test performance, a large body of empirical studies has investigated the relationship between individual characteristics and test performance from different perspectives. Kunnan (1998) summarized over 20 studies investigating different aspects

of test-taker characteristics, such as academic background, native language, culture, gender, and field dependence. He claimed that "this area of investigation has generated awareness among test developers and researchers that test-takers with different characteristics might be affected by a test or its items in ways that were not relevant to the abilities being tested" (p.6). Zeidner (1986, 1987), Kunnan (1995), and Chen and Hennings (1985) studied the influence of test-taker characteristics from the perspective of cultural background. Krashen (1985) studied the exposure to the target language. Alderson and Urquhart (1985), Chihara, Sakurai, and Oller (1989) studied students' academic background and linguistic proficiency. Other examples include learner variables in relation to ESL learning (Oller & Hinofotis, 1980) as well as the concept of an individual's attitude and motivation with regard to student achievement (Johnson & Krug, 1980). Taken together, these studies demonstrate that a variety of test-takers' characteristics influence test performance.

In addition, family socio-economic status (SES) has also long been documented to be related to students' performance in schools (Bianchi & Robinson, 1997; Fehrmann, 1987; Griffith, 1996, 1997; Hickman, Greenwood, & Miller, 1995; Lytton & Pyryt, 1998; Wendling & Cohen, 1980; White, 1982). Griffith (1996, 1997), and Lytton and Pyryt (1998) reported that SES was significantly related to academic achievement. White (1982), using meta-analysis techniques, examined almost 200 studies that considered the relationship between SES and academic achievement. He found that, although it was widely believed that SES was strongly correlated with measures of academic achievement, weak or moderate correlations were frequently reported. Fisher (2000), however, found that students with low SES had significantly higher GPAs than students from high SES backgrounds. He indicated that the SES status may not be as large a factor as it used to be.

The above opposing findings lend support to the importance of examining the relationship between SES status and ESL/ELD students' performance in the OSSLT context.

Additionally, test method has shown to be an important influence on test performance (Bachman & Palmer, 1982; Shohamy, 1983, 1984). The focus of this facet was on test format differences. Examples of similar studies which investigated different test formats and the relationship with test performance include Hancock (1994), Bennett, Rock, and Wang (1991), and Kobayashi (2002). Among these studies, the most heated debate is on the comparison of multiple-choice questions and constructed-response questions. Regarding this issue, there are two contrasting opinions. Some studies (e.g., Fitzgerald, 1978) found multiple-choice questions only required students' lower cognitive skills - the knowledge level. Others, however, held a different point of view. They found that it was not simply the test format that accounted for the degree of knowledge that was tested; it was how the test formats were constructed that made the difference (e.g., Hancock, 1994). Although the debate between the use of multiple-choice questions and constructed-response questions has provoked many theoretical and empirical studies, there have been few attempts so far to compare the performance differences between ESL/ELD students and non-ESL/ELD students on different test formats. Therefore, this study should contribute to a better understanding of ESL/ELD students' performance on the three reading test formats on the OSSLT as compared to their non-ESL/ELD counterparts. The writing component was not included in the examination of test format differences in this study, simply because in terms of writing, it is the writing tasks that differ, not the formats. As cited by Bachman (1990), test format refers to the nature of the expected response to the input.

Overview of the Thesis

This thesis is structured in five chapters. Chapter one, the present chapter, introduces the research context, the research questions, as well as the rationale for the research. Chapter two is a literature review that reports relevant studies conducted. After depicting the dataset used, the research designs and data analysis methods in Chapter three, results of the data analysis are presented in Chapter four. Chapter five discusses the results, the potential significance, and limitations and suggestions of the study.

Chapter Two: Literature Review

In this study, factors associated with ESL/ELD students' performance in the high-stakes literacy test in Ontario were examined, including factors from test-taker characteristics and test format differences. This chapter first briefly discusses the importance of literacy and presents some relevant national surveys on literacy development in Canada. The focus then shifts to ESL/ELD students' personal attributes that are associated with test performance, including both theoretical discussions and empirical studies. Literature about the relationship between test performance and test format differences is presented next. This review concludes with an illustrative figure summarizing the theoretical framework for this study.

Literacy Development in Canada

As discussed in Chapter one, the cultural and racial make-up of Canada has changed dramatically over the last decades due to immigration (Canada Immigration and Citizenship, 2002). The multi-cultural and multi-lingual features of the Canadian society make the skills to obtain English literacy for students from different backgrounds differ considerably. Montigny, Kelly, and Jones (1991) stated the significance of literacy skills in the following way:

> Literacy skills are basic and essential tools, which enable and enhance communication, understanding and awareness. Life-long learning skills allow individuals not only to develop professional skills but lead to a better understanding of the multiple facets of daily living in a complex environment. (p. 13)

In addition to the essential role of literacy in daily life, researchers have related the development of the whole nation's literacy level to the development of society. For example, the results of the Statistics Canada report (1996), "Reading the future: A portrait of literacy in Canada" demonstrate strong plausible links between literacy and the country's economic potential. Willms (1997) argues that literacy is an instrument of social power. "Literacy is so central to social and economic status, policy measures that decrease inequalities in literacy were fundamental to achieving tolerance, social cohesion, and equitable distribution of economic opportunities" (p. 6).

As important as literacy is in people's daily life as well as in its relationship with the development of society, interpreting literacy in the social and cultural context is anything but simple, direct, or unmediated (Graff, 1994). The nature of literacy is complex in at least two ways. First, literacy is viewed as a social phenomenon, which highlights the contingent nature of literacy. It is recognized that literacy requirements differ from one society, and one time, to another. Second, it is viewed as a cognitive skill, which relates literacy to background factors such as education and life experience. Its contingent nature and the cognitive aspect make literacy difficult yet important to assess, especially in a period of accelerated economic and social change (Montigny et al., 1991).

Montigny et al. (1991) used the data of the survey of Literacy Skills Used in Daily Activities (LSUDA) conducted by Statistics Canada in 1989. The objective of this survey was to provide a direct assessment of reading, numeracy, and writing skills of the Canadian adult population (16-69). The major findings of this study concerning reading skills revealed that only 62% of Canadian adults aged 16-69 had sufficient reading skills to deal with most everyday reading requirements. Further, the reading skills of 16% of

Canada's adults were too limited to allow them to deal with the majority of written materials encountered in everyday life.

Willms (1997) used the data from the International Adult Literacy Survey (IALS), a seven-country initiative conducted in 1994. He examined the relationships between literacy skills and family background and personal characteristics of Canadian youth aged 16 to 25, who were deemed to be "in transition" from completing their formal schooling to entering the labor market. The analyses in his study were based on a model which presumed that an individual's academic attainment, and ultimately occupational attainment, was largely determined by educational experiences and family origins. In this model, family origins had a direct effect on attainment through a variety of mechanisms that began at birth, as well as an indirect effect through education. For example, his model suggested that children from more advantaged backgrounds were more likely to have access to quality education, and greater financial and cultural capital to support educational activities during the elementary and secondary years. Hierarchical linear modeling (HLM) was used to examine the data that were nested within provinces. The results of Willms' study showed a relatively strong effect for family income and a moderate effect for parents' occupational status. The results also indicated that literacy was affected by experiences related to economic, cultural, and social capital. Most of the differences in literacy were attributable to respondents' socio-economic background, years of formal education, and whether their first language was the test language, which was a major issue in my study. In terms of the factor of whether their first language was the test language, Willms found that after five years of being exposed to the test language, most non-indigenous speakers might be expected to score close to the average of those whose first language was the test language, other things being equal.

As can be seen, many variables contribute to the development of literacy or school success, such as family income, parents' educational level, parents' occupational status, years of formal education, and the factor of whether the test language is the first language learned. As the major focus of the present study is on ESL/ELD students, a review of factors associated with ESL/ELD students' literacy acquisition and academic development is provided in the next section.

ESL/ELD Students' Test-takers Characteristics Associated with

Literacy Acquisition and Academic Development

The English as second language factor adds complexity to literacy acquisition and development picture. Not only do the above-mentioned factors such as family income and parents' educational level play roles in influencing students' academic performance, but other elements stemming from ESL /ELD backgrounds also exert influence. Literacy acquisition among second language learners involves a number of important factors (Ramirez, 1994). Reading and writing activities can be viewed from various perspectives. Seeing reading as an interactive process, for example, assumes that text comprehension is based on the notion that the reader constructs meaning on the basis of the text itself and on the prior knowledge that is available to the reader (Ramirez, 1994).

Reading and writing acts are depicted by Cummins (1980) in terms of cognitive demands and contextual features. Reading and writing activities occur as part of context-reduced communication that rely primarily on linguistic cues to establish meaning and in some cases involve suspending knowledge of the "real" world so as to interpret or manipulate the logic of communication correctly. Researchers have

recognized that contextual factors are critical in mediating cognitive processes in academic learning. In addition to the cognitive differences that test-takers bring to the test, affective variables also add to the differences when the interactions of their prior knowledge with these variables occur (Snow, 1993). Students' varying cultural and linguistic backgrounds provide them with different schemes or principles when they are taking tests. Since the complex nature of language and its relationship with culture are usually neglected in existing approaches to testing English-language learners (ELLs, a term used in USA; ESL/ELD students is a term used in Ontario), equitable and valid outcomes are not ensured (Solano-Flores & Trumbull, 2003).

In short, apart from the complex and contingent nature of literacy development itself, ESL/ELD students encounter more difficulties than their non-ESL/ELD counterparts, such as prior knowledge challenges, sociolinguistic differences and lack of knowledge of the real world in the target language environment (Ramirez, 1994; Cummins, 1981; Snow, 1993). Without considering these factors, we could not reasonably think about language and linguistic diversity (Solano-Flores & Trumbull, 2003). In addition, the conceptual and theoretical understandings of ESL/ELD students' challenges and problems have been studied empirically. Much of the empirical work has been done to document and analyze the relationship between ESL/ELD students and their individual variables. For example, Tuijnman (2001) investigated poor literacy proficiency among the second-language foreign-born population. Magiste (1984) analyzed the optimal age of students learning Swedish as their second language. The following is a brief review of some other important and relevant empirical studies.

Finnie and Meng's (2003) examination of visible minorities, aboriginal Canadians and whites took into account not only education levels but also immigration status,

literacy, and numeracy skill levels. Their result indicated that mothers' and fathers' education had strong influences on children's educational outcomes. For both men and women, having a degree generally had strong positive effects on both literacy and numeracy, while not speaking English as a first language, learning English or French later rather than earlier, and having experienced learning difficulties as a child had negative influences. Results also showed that immigrants had lower scores than the native-born, but their skills improved with their time in Canada. Moreover, Finnie and Meng (2003) reported that ethnic minority groups in Canada were extremely heterogeneous in terms of their income levels and measured literacy and numeracy skills. In particular, some minority groups had substantially lower levels of cognitive abilities than whites and other more economically successful minorities.

Montigny et al.'s (1991) survey used a questionnaire to profile the characteristics and the literacy skills of Canada's adult population. They set migration status as a variable in examining literacy skills and found that apart from schooling, a key factor related to literacy skills was the extent of knowledge and exposure to the literacy materials. The findings in their report showed that those who were born outside the country had significant differences when their reading skill profiles were compared to the Canadian-born population. They found that a consistent length and variety of exposure to printed materials in English affected the development of literacy skills. The age when English was learned and the length of residency in Canada were two vital measures that played roles in determining the exposure to everyday literacy materials.

Watt and Roessingh (1994) conducted a quantitative study tracking the educational progress of 232 ESL high school students in Alberta. Various factors were identified to determine which would be most useful as a basis for tracking ESL students' progress in

schools. Among these variables were gender, age on arrival, first language (L1) ability at entry, L1 educational status, country of origin, home language(s), numbers of languages spoken, educational gaps, immigration status, and English language proficiency at entry. The result showed that English language proficiency on entry into the school was reported to be the most powerful predictor of academic success.

Early (1992) used semi-structured interviews to identify factors that distinguished academically successful ESL students from academically less successful ESL students. Twenty-six ESL students were interviewed about the factors that might contribute to or accompany success in school. Four themes emerged as contributing factors to students' academic success: time, help, purpose and power, and study strategies. The theme of "time" was approached from many different perspectives: age on arrival, number of years of schooling, time spent on homework and studies after school. Help was a factor that contributed in two aspects: source of help and type of help sought. Successful students sought more help from teachers than from friends or family. Early (1992) included inner resources as the other source of help that ESL students might seek, which was represented in three ways: adaptability, courage, and confidence. The theme of purpose and power distinguished students in that the more successful students spoke of a goal they had set for the future, and more successful students felt in control of their circumstances and their destiny. In addition, academically successful students and academically unsuccessful students displayed different approaches to study strategies. The strategies which emerged as good discriminators were cumulative, disciplined hard extra hours of study.

Gardner, Polyzoi, and Rampaul (1996) conducted research examining the relationship between individual variables and ESL progress comparing nine Kurdish and seven Bosnian immigrants. They made the distinctions between alphabetic writing

systems and non-alphabetic (i.e. ideographic) writing systems, such as Chinese. They included variables that were quantifiable to determine which individual variable(s) best predicted the rate at which different learners acquired English. The variables they included were literacy level, ethnicity, gender, age, occupation, years of formal education, and other language(s) learned. In their study, interviews were conducted to examine other important predictors, such as aptitude for language learning, motivation and goals, confidence as a learner, previous or posttraumatic stress, family status, contact with native speakers of English, and current pressures or stress including family, health and financial concerns. T-tests, ANOVAs and simple linear regressions were conducted. They reported that multiple regression analysis was not employed because of the concern of multicollinearity and small sample size. An example of possible multicollinearity was that ethnicity might be confounded with years of formal schooling and other language learned. The results of the study showed that years of formal education in home countries was a significant predictor of academic success. The larger the number of years of schooling, the higher the language scores. Their conclusion supported the argument that well-developed literacy skills from the first language helped the acquisition and development of the second language. As with formal education in first language, students possessed a considerable amount of information about their first language system (e.g., how language functions), skills and learning strategies that allowed them to use language to acquire new information and met the demands of a classroom-based learning situation (e.g., using the dictionary, making notes, memorizing words or phrases), and a wide range of concepts on which to draw when confronted with new words in a second language. Those students with little or no formal education encountered comparatively greater

difficulty in adapting to a classroom-learning environment, as they did not have appropriate skills or strategies to which they could resort.

To summarize, in the field of education, determining the magnitude and causes of low achievement is a challenging task (Klesmer, 1994). In the cases of non-ESL/ELD students, the variables can be complicated. In the case of ESL/ELD students, more factors must be taken into consideration in order to draw a better understanding of the interacting factors that influence their academic achievements. According to Klesmer (1994), the additional factors would at least involve ESL/ELD students' socio-linguistic history, migration history, past schooling experiences in their native countries, and motivation and attitude towards their new culture.

The results of previous studies provide support for the investigation of literacy activities and language factors in relation to ESL/ELD students' literacy development. For example, hours and types of after-school reading and writing in English denote the time and types of exposure to the target language, which were evidentially shown to be important variables by Montigny et al. (1991) and Early (1992). The variables of English as the first language and home language(s) are phenomenal in predicting literacy levels. (Finnie & Meng, 2003; Gardner, et al., 1996; Montigny et al., 1991; Tuijnman, 2001; Watt and Roessingh, 1994; Willms, 1997)

Additionally, computer technology has ushered in a new era of mass media in students' learning (Wartella & Jennings, 2000). We could not talk about students' academic development without mentioning the influence of their computer use. This notion is supported by the idea that computer technology has become more and more accessible to students (Liu, Moore, Graham, & Lee, 2003). The increasing amount of time students are spending on computers at home raises questions of how the use of computer

technology might make a difference in their learning (Subrahmanyam, Kraut, Greenfield, & Gross, 2000). The exploration of students' home computer use after school as a variable in the current study intended to shed light on the relationship of literacy development and computer access in the current society.

In terms of the SES variable, the attention of this study was directed toward two measures: average household income and average household educational attainment, which were two vital elements in the traditional definition of socio-economic status (SES). Among various definitions of SES in the literature, the Michigan State Department of Education (1971) measured SES using the following three variables: (1) family income; (2) parents' educational level; and (3) parents' occupation. Some researchers claim that a strong relationship exists between all kinds of academic achievement variables and SES. This claim is widely accepted and cited as a self-evident fact. The higher the SES of a student's family, the higher his or her academic achievement (Griffith, 1997; Lytton & Pyryt, 1998; Wendling & Cohen, 1980). Others, however, find the relationship between SES and academic achievement difficult to support. The actual variation of various measures of the correlation between the two ranges from 0.1 to 0.8 (White, 1982). Instead of making SES a convenient label to attach to a variety of different combinations of variables, the present study examined two indicators: average household income and average household educational attainment.

<center>Test Format Differences and Test Performance</center>

In addition to the test-taker characteristics, test format is also associated with students' performance (Bachman, 1990, 2000). The facet of expected responses in Bachman's (1990) category of test methods has three subcategories: type of expected

response (selected vs. constructed); form of expected response (language, non-language, or both); and language of expected response (native, target, or both) (Bachman, 1990). Among these three subcategories, the facet of type of expected response is crucial and has been the subject of many debates because it is generally assumed that different test formats elicit different levels of skills. In particular, many researchers have compared the differences between multiple-choice format and constructed-response format (Bennett et al., 1991; Fitzgerald, 1978; Gay, 1980; Hancock, 1994; Lukhele, Thissen, & Wainer, 1994; Riley, 1996; Shohamy, 1984). In this study, I consider "test format" equivalent to Bachman's (1990) term of "type of expected response." The following literature centers on the studies conducted comparing different test formats. Special attention is paid to the comparison between multiple-choice questions and constructed-response questions. As on the OSSLT, it is these formats that constitute the reading component.

Multiple-choice Questions vs. Constructed-response Questions

A number of studies have compared multiple-choice questions and constructed-response questions both theoretically and empirically. As noted in Hancock (1994), multiple-choice (MC) test items were first implemented in 1937 and have gained wide popularity in almost all kinds of standardized testing. Lukhele et al. (1994) defined the term "constructed-response" as any question format that required the test-takers to produce a response in any way other than selecting from a list of alternative answers.

Researchers have claimed that there are advantages in MC items, e.g., its scoring efficiency and freedom from ambiguity (Gay, 1980); its being "economically practical" and allowing "reliable, objective scoring" (Wainer & Thissen, 1993, p. 103). However, there are studies that argue the incompetence of MC questions. For example, MC has

been challenged as being inadequate to fully assess the dimensions of cognitive

performance, as it has limited opportunity for demonstrating in-depth knowledge afforded

by this format. MC questions "may emphasize recall rather than generation of answers"

(Wainer & Thissen, 1993, p. 103). In addition, there is the possibility that test-wiseness

can contaminate the measurement. "Test-wiseness" includes a variety of general

strategies related to efficient test taking (Bachman, 1990). As far as multiple-choice

questions are concerned, ruling out as many alternatives as possible and then guessing

among the ones remaining is considered to be an example.

Constructed-response formats are favored by some researchers because of the idea

that they can measure traits that can not be tapped by multiple-choice items, for example,

assessing dynamic cognitive processes (Bennett, Ward, Rock, & Lahart, 1990).

Constructed-response questions are also thought to replicate more faithfully the tasks

test-takers face in academic and work settings. Furthermore, constructed-response

questions are considered to provide tasks that "may have more systemic validity" (Wainer

& Thissen, 1993, p. 103). One well-cited theoretical rudimentary work was laid by Bloom,

Englehart, Furst, Hill, and Krathwohl half a century ago (1956). The six categories of

Bloom's taxonomy are knowledge, comprehension, application, analysis, synthesis, and

evaluation. Hancock (1994) indicated that the understanding of the taxonomic levels and

the linearity of those levels were of direct importance.

Opponents of multiple-choice test format claim that it is only knowledge-level that

can be touched upon, and even those proponents of multiple-choice questions admit that

the synthesis level in Bloom's taxonomy can not be reached by multiple-choice questions,

as synthesis level, as the name implies, requires test-takers to demonstrate their abilities

to integrate the information on the test with their own understandings and create a unique

product by them. In other words, constructed-response questions require students to combine their reading skills with writing skills.

Because constructed-response questions require test-takers to construct their own answers, the assumption is that this format must involve higher-level thinking. But this idea is challenged by some researchers. For example, Hancock (1994) investigated the comparability with which the multiple-choice and constructed-response formats assessed particular levels of complexity within the cognitive domain. He constructed exams for two measurement classes with half multiple-choice and half constructed-response items. Equal numbers of items in each format were written by using the first four levels of Bloom's taxonomy (knowledge, comprehension, application, and analysis levels). His argument was that given sound test construction, multiple-choice items were able to measure the same abilities as constructed-response items. The results indicated a pattern of high disattenuated correlations between multiple-choice and constructed-response questions. Hancock (1994) implied that to ensure that multiple-choice questions tap higher cognitive levels, skills are required on the part of the test constructors to develop distrators that ensure desired cognitive levels.

Based on the theoretical discussions, some empirical studies have been conducted to examine the effects of different test formats on students' performance. For example, Kobayashi (2002) investigated the relationship between the students' test performance and the two other variables: text types and test formats. She tested 754 college EFL (English as Foreign Language) students in Japan on four types of rhetorical organization: association, description, causation, and problem-solution, ranging in the degree of interconnectedness of ideas, from loosely organized to tightly organized. Three test formats were employed: cloze, open-ended questions, and summary writing. Her findings

were that both text types and test formats had significant impact on the EFL students' performance. Learners of different proficiency levels were affected in different ways. Learners with higher English language ability were more susceptible to different test formats. The results demonstrated that different test formats and different types of items within the same format, measured different aspects of reading comprehension. Her findings also supported the concept of a "linguistic threshold" (p. 210), in which learners below a certain level of proficiency had difficulty understanding beyond sentence-level or literal understanding. Higher proficiency learners, on the contrary, were more aware of overall text organization.

Shohamy (1984) investigated the effect of different testing methods, levels of reading proficiency and languages of assessment on second language (L2) reading comprehension of readers of English as a foreign language. Using multiple-choice and open-ended questions presented in both the subjects' first language (L1) and L2, she found that learners performed better on multiple-choice questions presented in their L1 and these effects were greater for students with low levels of reading proficiency. Her findings indicated significant effects existed for all three variables - testing method, text, and language - on students' scores in reading comprehension.

Riley and Lee (1996) compared two global test formats - summary and recall protocol on the reading performance of two levels of early-stage L2 readers of French. They asked half of the subjects to read a passage and to write a summary of the passage, and asked the other half to read the passage and recall it. The performance was compared on the total numbers of idea units, the numbers of main ideas, and supporting ideas and details. Findings indicated a significant difference in the performance on the two tasks. The summaries contained a significantly higher percentage of main ideas than details

whereas the recall protocols contained a significantly higher percentage of details than main ideas. The results indicated that summarization techniques enabled students to interact with the text at a more global level and should therefore be encouraged in instruction.

Fitzgerald (1978) investigated the differential performance of students at three grade levels in two cultures (American and Irish) using three test formats: multiple-choice cloze, maze, and cloze[2]. The results indicated that students from the two different countries produced significantly different scores at grade 3 and grade 4. Concerning the levels of difficulty for test formats, the study supported the assumption that MC questions would produce the highest student scores since it was a recognition task and all the options were syntactically plausible. This result was upheld for both cultural groups. However, some differences were recognized within the two cultures. For example, the integrated program with considerable creative writing in Ireland accounted for the relatively higher scores on cloze for Irish. On the contrary, the higher scores on multiple-choice and maze for students from United States were explained by their more skill oriented and less integrated program in United States. Thus it was concluded that the differential performance between two cultural groups reflected characteristics of the educational programs within the cultures, such as their different foci of orientations in developing reading skills. Another contribution of this study was that it not only tried to examine the relationship of student performance between two cultural groups on the three

---

[2] In this study, both multiple-choice and maze formats represented a recognition task, while the cloze format represented a completion task. The difference between multiple-choice and maze was that the choices for multiple-choice questions were all syntactically plausible, while the maze had one choice syntactically plausible but semantically implausible; another choice both syntactically and semantically implausible.

test formats, but it also intended to correlate performance with other external variables, such as sex, age, personality and achievement.

Summary

To reiterate, previous studies have demonstrated that students' test performance is not solely related to their language abilities. Various test-taker characteristics, which may not be part of students' language abilities, are important to study in relation to students' test performance. Test formats, which elicit students' different levels of cognitive abilities, are another important area to be examined closely. Considering the high stakes of the OSSLT and the heterogonous nature of its test-takers, it is, therefore, of great importance to look into test-taker characteristics and the test itself. By combining the literature reviewed so far, the following research framework is presented to guide this study.

Figure 1

Theoretical Framework of This Study

October 2003 OSSLT Test-takers
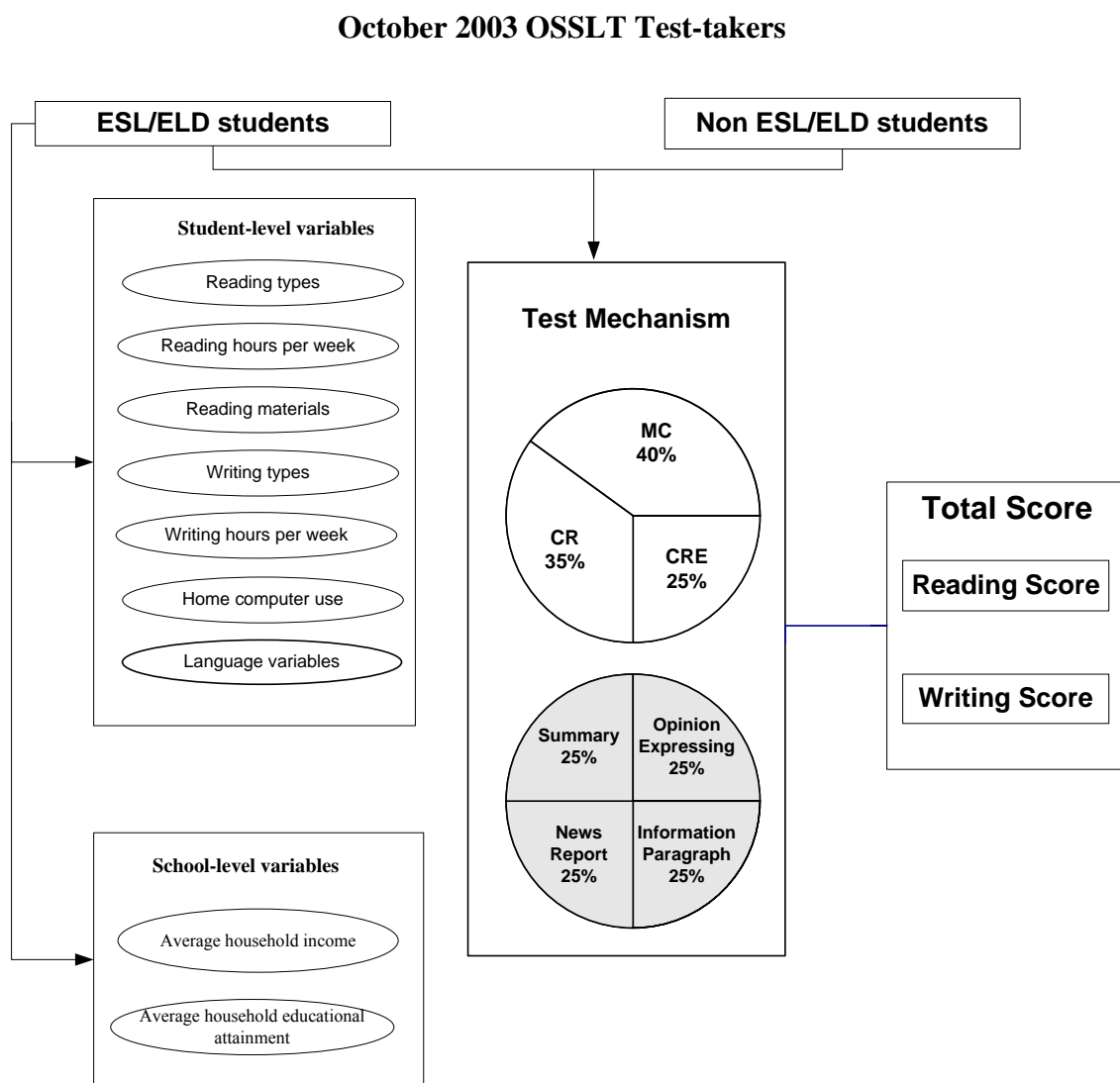
The above framework guides the three research questions in the present study. The

OSSLT test-takers in October 2003 were roughly divided into two major groups of

students: ESL/ELD students and non-ESL/ELD students. To address the first two

research questions, that is, to explore the relationship between ESL/ELD students' test

performance and their individual test-taker characteristics, to explore the relationship

between school test performance and average school-level socio-economic status, only

ESL/ELD students were examined. To address the third research question, that is, to compare ESL/ELD students and non-ESL/ELD students on the reading test formats, both groups were examined. The column of the test mechanism depicts the two components of the OSSLT. The reading component has three different test formats, which leads to a reading total score. The writing component has four writing tasks, which leads to a writing total score. Stated differently, the aim of the study was to examine closely to what extent the outcome of the test, the test scores, were associated with or influenced by the test-taker characteristics and the test mechanism of the OSSLT.

Due to the availability and nature of the data in this study (which are fully discussed in the next chapter), the concept of test-taker characteristics examined in this study was composed of two levels: (1) student level test-taker characteristics, which mostly referred to the characteristics of students' after-school literacy activities, home computer use, and home language background; (2) average school-level socio-economic indicators, which included annual income indicator and parental education indicator. Owing to the fact that these test-taker characteristics were collected at the student level and the school level respectively, they were examined in relation to different levels of achievement data: individual achievement data and aggregated school-level achievement data. More detailed descriptions of the data and variables included in the study are provided in the next chapter.

Chapter Three: Methods

This chapter delineates the nature of the data used and the methods employed in addressing the research purposes of this study. The three research purposes were: 1) to examine the relationship between ESL/ELD students' characteristics and their OSSLT performance; 2) to examine the relationship between these students' aggregated test performance and the school-level socio-economic status; 3) to investigate any possible features emerging from the test performance of ESL/ELD students in the three test formats in the reading part of the OSSLT in comparison to their non-ESL/ELD counterparts.

Description of the Dataset Used in the Study

Three sources of data were used in order to fulfill the research purposes. They were the October 2003 OSSLT data; the 2003 EQAO Student Questionnaire data; and school-level SES data from the 2002 Education Quality Indicator Program (EQUIP). A detailed depiction of the nature of the data is provided in the following sections, including the delineation of the test data, the questionnaire context, and the procedures of merging and aggregating the data to arrive at what was used in the current study.

Test Data

For the 2003 test data, four types of test-takers were included by EQAO: 1) first time eligible students who completed both the reading and the writing tests; 2) previously eligible students who were required to take both the reading and the writing. They were the students who either failed both in previous tests, or deferred previous tests (February

2002, October, 2002); 3) previously eligible students who were required to take the reading only, as they only failed the reading part in previous tests; 4) previously eligible students who were required to take the writing only, as they only failed the writing part in previous tests. Except for the first type of the students, the rest were considered as test re-takers, who either deferred from previous test administrations or failed at least one part of the previous tests. Those who deferred from this October 2003 test or had no record on this test were excluded from the sample.

Three sets of October 2003 test data were obtained from EQAO, the first set was ESL/ELD students' test data, in which there were 4,311 ESL/ELD students (public and private schools) who wrote the test in October 2003. The second and third sets were test data of non-ESL/ELD students, with 5,000 non-ESL/ELD students who passed the test, and 5,000 non-ESL/ELD students who failed the test. A stratified sample of non-ESL/ELD students who either passed or failed the test were selected in conformity with the overall pass-fail ratio in October 2003 (23% fail, 77% pass). Therefore, to better represent the overall pattern, 77% of the non-ESL/ELD students who passed (3,834 cases), and 23% of the non-ESL/ELD students (1,169 cases) who failed were selected. A total of 5,003 non-ESL/ELD students' test data were compared with 4,311 ESL/ELD students' test data to explore if any differential performance between these students on the reading sections, and how these differences, if any, manifested themselves in terms of performance patterns.

Questionnaire Context

Each student who wrote this secondary school literacy test was asked to complete a questionnaire at the time of the test administration (see Appendix I). The questionnaire

was composed of four major parts: after-school reading activities, after-school writing activities, home computer use, and home language background. The data were measured on either a dichotomous scale or an ordinal scale (see Table 1). For example, a dichotomous question required students to indicate the types of materials they read in English outside school. They were asked to choose whether they read non-fiction, books, comics, or magazines after school. An ordinal question asked students to choose from the four-point scale of the hours they devoted to their after-school reading or writing: *less than one hour; from one to three hours; from three to five hours;* and *more than five hours*.

Table 1

Questionnaire Data Description

	Major Question Categories	*Variables Involved*	*Measurement Scale*
I	*Reading Types*	1) Non-fiction 2) Comics 3) Internet 4) Letters 5) Magazines 6) Manuals 7) Newspapers 8) Novels 9) Poetry 10) Religious	Dichotomous
	Reading Hours Per Week	11) Number of reading hours 　a) one hour or less 　b) more than one hour and 　　less than three hours 　c) more than three hours and 　　less than five hours 　d) five hours or more	4-point scale
	Reading Materials Possessed at Home	12) Dictionary 13) Books 14) Newspapers 15) Magazines	Dichotomous
II	*Writing Types*	16) E-mail 17) Letters 18) Notes 19) Songs 20) Stories 21) Work-related writing	Dichotomous
	Writing Hours Per Week	22) Number of writing hours 　a) one hour or less 　b) more than one hour and 　　less than three hours 　c) more than three hours and 　　less than five hours 　d) five hours or more	4-point scale
III	*Number of Home Computer Use*	23) Number of computer use at 　home for school work 　a) I don't have a computer at 　　home 　b) I never or hardly ever use 　　the computer for school	5-point scale

		work	
		c) I use computer once or twice a month for school work	
		d) I use the computer once or twice a week for school work	
		e) I use the computer almost every day for school work	
IV	*Language Background*	1) English is the first language learned	Dichotomous
		2) Language spoken at home a) only or mostly English b) another language (or languages) as often as English c) only or mostly another language (or other languages)	3-point scale

In brief, each dichotomous item was treated as an individual question in the descriptive analysis in this study. Therefore, concerning after-school literacy activities, the first three parts of the questionnaire were included: part I reading activities, part II writing activities, and part III home computer use. 19 out of the total 23 items in these three parts were collected on a dichotomous scale (see Table 1), which sets limit to the exploration of the relationship. In addition, two questions concerning students' language background were elicited in the fourth part of the questionnaire. In order to explore the relationship between ESL/ELD students' individual characteristics and their literacy performance, 4,311 ESL/ELD students' questionnaire data were linked with their test data using the student identity numbers.

EQUIP Data

The school-level data employed in this study were provided by the Education Quality Indicator Program (EQUIP), which were designed to measure and describe a range of significant features of Ontario education at the school, school board and provincial levels. It collected and reported information about the many factors inside and outside the school community that affected student achievement across Ontario. The EQUIP intended to provide detailed information on the unique and complex characteristics of schools, school boards and educational systems. Research showed that many factors influenced student learning, some of which lay outside schools. EQUIP also provided a context for examining and understanding student achievement scores. Contextual features, as defined in the EQUIP framework, included "parental involvement in school, students' linguistic background, and community socio-economic status" (EQUIP: The framework, 2004, p.1). Test scores were placed in a broader socio-economic context, encompassing the teaching and learning environment, student mobility and the accessibility of computers and resource materials (EQUIP: The framework, 2004).

EQUIP data was compiled from a number of existing sources, including the Ministry of Education, EQAO assessments and questionnaires, and Statistics Canada. Eight themes were developed, one of which was the theme of *community, student and family demographics*. Socio-economic status was gathered at provincial, school board, and school levels. In this study, school-level SES information in the EQUIP data was used to predict the school-level literacy performance.

It should be noted that EQUIP data was not collected every year. The 2002 data was the most recent one, which was from the 2001 census. The justification of the use of the 2002 school-level SES data was that compared with other factors, the school SES data did not vary very much in the intervening years. Information used from the EQUIP data mainly fell into two SES indicators: household income levels and parental educational achievement levels (see Table 2).

Table 2

EQUIP Data Used: School-Level SES Indicators

	Major SES Indicators	*Items Involved*	*Data Property*
I	*Annual Household Income Levels*	1) <$20,000 2) $20,000-$39,999 3) $40,000-$59,999 4) $60,000-$79,999 5) $80,000-$99,999 6) above $100,000	In Percentage
II	*Parent Educational Levels*	1) <grade 9 2) >grade 9 but incomplete high school 3) complete high school 4) some post-secondary 5) at least bachelor's degree	In Percentage

From the above table, it was found that both SES indicators were measured in percentages. Percentages of annual household income levels were collected at six levels, and each school had all these six levels of annual household income percentages reported in the data. For example, one school might have 21% of its average annual household income at lower than $20,000 level; another school might have 18% at that level. Meanwhile, percentages of parental educational levels were collected at five levels, and each school had all these five levels of parental educational percentages reported in the

data. For example, one school might have 19% of its parental education at the *lower than grade 9* level, and another school might have 15% at that level. The methods that I employed to use the above data are presented in the next section.

Research Designs and Data Analyses Methods

The following sections describe the design of each research question. Methods employed in answering the three research questions are presented in sequence.

Methods Employed in Answering Research Question One

First of all, a descriptive analysis of all the variables included in the EQAO Student Questionnaire was performed for the purpose of obtaining an overall picture of ESL/ELD students' involvement in different literacy activities. Then an exploratory factor analysis was performed in order to reduce the number of variables related to the 23 items of after-school literacy activities. The aim of the exploratory factor analysis was to explore the potential constructs of after-school literacy activities examined in the questionnaire. The variables concerning students' home language patterns and language background were examined separately.

Principal component factoring and the varimax method were used in the factor analysis because they maximize variance and help the interpretation of constructs deduced. Principle component factoring, according to Gorsuch (1983), "gives the best least-squares fit to the entire correlation matrix, and each succeeding factor accounts for the maximum amount of the total correlation matrix obtainable" (p. 99). The varimax solution seeks "the rotation position where the variance is maximized across all factors in the matrix" (Gorsuch, 1983, p.185). Furthermore, given the arbitrary nature of the

question of how many factors to extract, two different methods were attempted in order to explore and ensure meaningful interpretability of the solutions and compare the relative model fitness and construct representation. The two methods were the Kaiser Criterion and the scree plot test. The Kaiser Criteria retains factors with eigenvalues greater than 1 (Crocker & Algina, 1986). The Scree plot test looks at "predominant factors [that] accounts for most of the variance and are large, whereas the other factors are quite numerous but small" and "the point where the factors curve above the straight line formed by the smaller roots gives the number of factors" (Gorsuch, 1983, p. 167). Factor loadings that were larger than 0.3 were reported, as this cutoff is conventionally regarded as a "meaningful loading" to be reported (Pedhazur & Schmelkin, 1991, p. 603). The ratio of the squared factor loadings for a given variable shows the relative importance of the different factors in explaining the variance of the given variable. Thus with factor loadings greater than 0.3, the variance explained by the variables would be equal to or greater than 9%. This means that factor loadings lower than .3 were deleted without being displayed in the result table and were not counted toward any factors. Meaningful interpretations were sought when double loadings occurred.

Then the corresponding reliability estimates were obtained. Multiple regression analyses of the factors extracted were conducted to find out how well these literacy activities factors predicted the ESL/ELD students' performance on the OSSLT. The method chosen was stepwise regression. Stepwise regression begins with an empty model and adds variables in order of importance to prediction. Variables are removed if their contributions are no longer significant, after other variables are added (Glass, 1996). These procedures are best for situations in which one is interested in developing a good prediction equation but not in testing theories. They also may be useful in exploratory

research, which will result in models to be considered theoretically and tested more carefully later. Hence the stepwise method was adopted as it suited the nature of the present study. To explore how test-takers' factors predicted students' test performance, test scores were regressed on all the factors extracted in the factor analysis. Separate regression analyses were performed for the reading total score and the writing total score to see if any different patterns regarding these two literacy aspects would emerge. The significance level in this study was set to be α <.01.

Methods Employed in Answering Research Question Two

For the purpose of investigating the relationship between the aforementioned average school-level SES indicators and ESL/ELD students' literacy performance on the OSSLT, the existing test data and the EQUIP data were merged using the school identity codes. As stated in the previous sections of the data descriptions, the 2003 test data were collected at the individual level, and 2002 EQUIP data were collected at the school level. Therefore, the 2003 student test data were aggregated by school identity number (sch_id). All necessary variables to be included in the new data set were selected as *aggregate variables*, in this case, all the test score data. Hence students from the same school were aggregated, and the mean values of the aggregated data were used. Steps were then taken to merge the aggregated test data with the EQUIP data according to the school identity codes. Of the 117 schools were examined, the number of ESL/ELD students spanned a considerable range from a low of 15 to a high of 136. The new merged file was thus created to enable further analyses.

In order to carry out the investigation of the relationship between school-level SES indicators and aggregated school performance, the following calculations were completed.

First, an overall income index for each school was obtained by assigning different

weights to the relevant income variables listed in Table 2. The purpose of obtaining an

overall income index[3] was to get an income indictor that combined the information given

in the original data. Similar procedures were followed to obtain an overall parental

education index[4] for each school (See Appendix II for the EQUIP SES variable view). In

addition, before multiple regression analyses were conducted to investigate the

relationship between school-level SES indicators and the aggregated school performance,

collinearity between the education index and the income index were checked. A tolerance

value was obtained to determine if these two SES indicators could be regarded as two

independent variables.

Methods Employed in Answering Research Question Three

With regard to the comparison between ESL/ELD students' performance and

non-ESL/ELD students' performance in the three reading test formats, three format level

test data on OSSLT were required from both groups of ESL/ELD students and

non-ESL/ELD students: MC scores, CR scores and CRE scores.

The initial hypothesis was that there would be a greater performance gap between

ESL/ELD students and non-ESL/ELD students in MC questions than in CR and CRE

questions. Because compared with MC questions, which might only require students'

knowledge-level ability, CR and CRE questions require not only knowledge-level ability

---

[3] The formula to obtain the income index
   COMPUTE income = [(p_be_20 * 1) + (p_20_39 * 2) + (p_40_59 * 3) + (p_60_79 * 4) +
   (p_80_99 * 5) + (p_g_100*6)]/1
[4] The formula to calculate the education index
   COMPUTE education = [(p_b_g9 * 1) + (p_g9_bhs * 2) + (p_hi_sc * 3) + (p_ps_sec * 4) + (p_bach
   * 5)]/1

but synthesis-level ability. Also, CR and CRE questions required students to integrate personal knowledge and experience to extend meaning in reading, which would most possibly place ESL/ELD students at a disadvantage.

Necessary actions were taken to allow statistical analyses to be executed. First, I merged the 4,311 ESL/ELD students' files with the 5,003 non-ESL/ELD students' files. Before merging, a new variable "eslstatu" was created to differentiate the two groups. I set 1 as the ESL/ELD group, and 2 as the non-ESL/ELD group. The sample size of the merged file was 9,314.

To compare how ESL/ELD students and non-ESL/ELD students performed on each test format, descriptive statistics of raw scores were run. At the same time, other indicators, e.g., standard deviation, skewness, and kurtosis were obtained. Discriminant analyses were then performed to see which format(s) could be used to distinguish the ESL/ELD group and their counterparts. Discriminant analysis is the statistical technique most commonly used to classify cases into two or more groups based on various characteristics of cases and to predict group membership for new cases whose group membership is undetermined (Norusis, 1988). In the two-group case, discriminant function analysis is analogous to multiple regression. Discriminant analysis has two steps: first, an F test (Wilks' lambda) is used to test if the discriminant model as a whole is significant; second, if the F test shows significance, then the individual independent variables are assessed to see which differ significantly in mean by group and these are used to classify the dependent variable (Garson, 2005).

The discriminant equation is: $D = a + b_1X_1 + b_2X_2 + \ldots b_iX_i$ in which X_i represents each independent variable, b_i represents the corresponding coefficients estimated from the data, and the D is the predicted group membership. The resulting coefficients provide the

maximum separation among the groups. In this case, the independent variables were MC scores, CR scores, and CRE scores and D was predicted ESL/ELD membership.

Follow up correlation analysis was conducted to check if multicollinearity among the three test formats was a concern. Finally, classification results were obtained to demonstrate how the discriminant functions work to differentiate the ESL/ELD students from non-ESL/ELD students.

Summary

In this study, three sources of data were used to: 1) predict ESL/ELD students' OSSLT performance, using individual test-takers characteristics on the EQAO student questionnaire; 2) predict school-level ESL/ELD students' OSSLT performance, using aggregated socio-economic status indicators; and 3) answer whether differential performance on test formats exists between ESL/ELD students and non-ESL/ELD students. Procedures were taken to aggregate and merge different data from different sources to allow the desired analyses to be performed. Statistical analyses, including factor analysis, stepwise multiple regression, and discriminant analysis were completed to address the three research purposes. The results of these analyses are presented in the following chapter.

Chapter Four: Results

This chapter has three sections. Section one presents the results of the analyses conducted to investigate the relationship between ESL/ELD students' OSSLT performance and their characteristics, which included their after-school literacy activities and home language patterns. Section two contains the results of the analyses conducted to investigate the relationship between these students' aggregated test performance and the average school-level socio-economic status. Section three contains the results of the analyses conducted to address the inquiry of the test performance in different reading formats between the ESL/ELD students and the non-ESL/ELD students.

Student-level Test-taker Characteristics and Their OSSLT Performance

Descriptive Analysis

Descriptive analyses of all the variables in the EQAO Student Questionnaire were performed for the ESL/ELD students. The questionnaire examined two major aspects of the students' information, after-school literacy activities and home language patterns. The descriptive results are presented in the following three tables. Table 3 contains the descriptive statistics of the 20 dichotomous variables of after-school literacy activities in which the mean of each item represents the percent of the students' participation in each activity. Table 4 contains the frequency of the three literacy-related non-dichotomous variables. Finally, Table 5 contains the frequency of the two home language background variables.

Table 3

Descriptive Statistics of Dichotomous After-school Literacy Activities Variables (*N*=4278)

	Questionnaire items[5]	%
I Reading	**Types of reading in English outside school most weeks**	
	Non-fiction	19.71%
	Comics	27.12%
	Internet	68.91%
	Letters	26.83%
	Magazines	51.29%
	Manuals	13.51%
	Newspapers	43.20%
	Novels	51.36%
	Poetry	30.88%
	Religious	15.33%
	English-language reading materials at home	
	Dictionaries	75.41%
	Books	73.40%
	Newspapers	56.87%
	Magazines	51.85%
II Writing	**Types of writing in English outside school most weeks**	
	Email	80.93%
	Letters	32.66%
	Notes	28.42%
	Songs	23.07%
	Stories	17.93%
	Work-related	33.38%

---

[5] See Appendix I for the actual questions asked in the questionnaire.

The first horizontal panel in Table 3 contains the results of the dichotomous survey items related to reading in English. The most common reading form for ESL/ELD students was through the internet (68.91%). Just over half of the students reported that they read novels or magazines (51.36%, 51.29%). Less than half of the students read the other types of reading materials during most weeks. Manuals and religious reading were the least reported forms (13.51% and 15.33%). Students were also asked to identify the English language reading materials they possessed at home. All of the four types listed in the questionnaire were reported to be owned by the majority of the students (ranging from 51.85% to 75.41%). The second horizontal panel contains the results of those dichotomous survey items related to writing in English. Other than writing emails, which 80.93% of the ESL/ELD students reported that they did, about one third of the students reported that they did letter writing (32.66%) or work-related writing after school (33.38%). The other forms of writing, e.g., notes, songs, stories, were even less reported, ranging from 17.93% to 28.42%. Based on these results, it appears that ESL/ELD students were more likely to be involved in reading activities as opposed to writing activities outside school.

The results in Table 4 provide the hours that these students engaged in literacy activities in English outside school, excluding homework hours. The response rates to the three listed questions varied a little. This variability was ignored given the comparatively large sample size. It was found that ESL/ELD students spent relatively more time in reading than writing. Most students spent less than 3 hours per week in English-related after-school literacy activities (72.75% for reading and 78.70% for writing). Computers were reported to be frequently used at home for school work. A total of 39.03% of the students used computer once or twice a week for school work, and 37.13% of the students

reported that they used computers almost everyday for school work. Less than one tenth

of the students did not have a computer at home or hardly ever used computer for their

school work after school. The results indicated that the hours students devoted to literacy

activities outside school in English were limited. The results also indicated that many

students utilized computers frequently to help them complete their school work.

Table 4

Frequency of Non-dichotomous After-school Literacy Activities Variables

Reading hours per week in English outside school (homework exclusive) (N= 4028)		Writing hours per week in English outside school (homework exclusive) (N= 4154)		Computer use at home for school work (N=4140)	
Category	%	Category	%	Category	%
1 hour or less	30.38%	1 hour or less	38.40%	Don't have a computer at home	5.94%
More than 1 hour and less than 3 hours	42.37%	More than 1 hour and less than 3 hours	40.30%	Never or hardly ever use the computer for school work	3.96%
More than 3 and less than 5 hours	17.23%	More than 3 and less than 5 hours	13.38%	Use the computer once or twice a month for school work	13.94%
5 hours or more	10.01%	5 hours or more	7.92%	Use the computer once or twice a week for school work.	39.03%
				Use the computer almost every day for school work.	37.13%

Not surprisingly, in terms of first language, the majority (93.18%) of the students spoke English as their second language (see Table 5). A small percentage (6.82%) reported English as their first language. In terms of language(s) spoken at home, the results revealed that 5.98% of the students only or mostly spoke English, 34.45% of the students spoke another language as often as English, and 59.57% of the students only or mostly spoke another language. Tentative classifications of these students were attempted according to the definitions of ESL and ELD students offered by Ministry of Education and Training and EQAO. Combining the results from the two home language variables above, it indicated that the majority of these test-takers were ESL students. And they belonged to two categories. 59.57% of them mainly spoke another language at home. 34.45% of them spoke a combination of English and their native language at home. In contrast, within this sample, those students reporting English as first language were considered to be ELD students. Hence they mainly spoke English at home, even though it was possible that they may speak other varieties of English (Ministry of Education and Training, 1999).

Table 5

Frequency of Home Language Variables

First language n= 4178		Language spoken at home n= 4165	
Category	%	Category	%
English	6.82%	Only or mostly English	5.98%
Not English	93.18%	Another language (or languages) as often as English	34.45%
		Only or mostly another language (or other languages)	59.57%

In brief, the descriptive analyses above provided some general patterns of ESL/ELD students' after-school literacy activities and their home language features. In order to further explore the potential constructs of ESL/ELD students' after-school literacy activities, exploratory factor analyses were performed.

Factor Analysis

Using the procedures outlined in the Methods Chapter, the 23 literacy-related items in the survey (dichotomous and non-dichotomous inclusive) were factor analyzed. Two methods were attempted: the Kaiser Criteria method and the scree plot method. The use of the Kaiser Criteria method resulted in a 7-factor model while the scree plot method indicated a 3-factor model. After a comparison of the two models, only the results from the Kaiser Criteria method were presented and discussed, as this model was readily comprehensible and met the criteria of interpretability. The output of the scree plot is, however, included in the Appendix III for the reference purpose.

Table 6 presents the 7-factor solution with the factor loadings based on the rotated component matrix. Eigenvalues, individual variance explained, and accumulated variance explained are presented. The items with common factor loadings were then used to determine the label for each factor. Factor 1, newspaper and magazine literacy activities, contained 4 variables: item 5, types of reading 'magazines'; item 7, types of reading 'newspapers'; item 14, reading material at home 'newspapers'; and item 15, reading material at home 'magazines'. Factor 2, non-fiction literacy activities, contained 5 variables: item 1, reading non fiction; item 6, reading manuals; item 10, religious reading; item 18, writing notes; and item 21, writing work-related tasks. Factor 3, e-literacy activities, contained 3 variables: item 3, reading through internet; item 16, writing emails;

and item 23, frequency of home computer use for school work. Factor 4, song and poetry literacy activities, contained 2 variables: item 9, reading poetry and item 19, writing songs. Factor 5, literature literacy activities, contained 4 variables: item 8, reading novels; item 20, writing stories; item 13, having books at home; and item 12 having dictionary at home. Factor 6, literacy hours, contained 2 variables: item 11, reading hours per week; and item 22, writing hours per week. Factor 7, letter reading and writing, contained 2 variables: item 4 and item 17: letter reading, and letter writing.

Table 6

7-factor Rotated Component Matrix

Item		Factor						
		1	2	3	4	5	6	7
5	Types Of Reading: Magazines	.71						
7	Types Of Reading: Newspapers	.44						
14	Reading Material: Newspapers	.61						
15	Reading Material: Magazines	.75						
1	Types Of Reading: Non-Fiction		.52					
6	Types Of Reading: Manuals		.65					
10	Types Of Reading: Religious		.51					
18	Types Of Writing: Notes		.48					
21	Types Of Writing: Work		.57					
3	Types Of Reading: Internet			.75				
16	Types Of Writing: Email			.79				
23	Home Computer Use			.48				
19	Types Of Writing: Songs				.80			
9	Types Of Reading: Poetry				.76			
8	Types Of Reading: Novels					.76		
20	Types Of Writing: Stories					.51		
13	Reading Material: Books					.55		
12	Reading Material: Dictionary			.38		.35		
11	Hours Reading Per Week						.80	
22	Hours Writing Per Week						.83	
4	Types Of Reading: Letters							.71
17	Types Of Writing: Letters							.76
	Eigenvalue	3.88	1.65	1.52	1.31	1.21	1.08	1.02
	Variance explained (%)	8.43	8.16	7.51	6.89	6.71	6.66	6.29
	Accumulated variance explained (%)	8.43	16.59	24.09	30.98	37.69	44.35	50.65

As noted in the above table, these seven factors accounted for more than half of the total variance (50.65%). The amount of variance explained by these factors were close, ranging from the highest of 8.43% by factor 1 newspaper and magazine literacy activities, to the lowest of 6.29% by factor 7 letter reading and writing. However, within this matrix, item 12, having dictionaries at home, had double loadings slightly higher than .3. Since

the inclusion of this variable in the factor of e-literacy activities did not make much sense, it was included in the literature literacy activities factor. Moreover, item 2, reading comics, did not have any loadings larger than .3 (See Appendix III), and was thus discarded from further analysis. Overall, 22 variables were extracted into 7 major factors to represent the constructs of these students' after-school literacy activities. Pearson correlations were computed for these factors (see Appendix IV). The correlations among the 7 factors were all significant albeit low (0.07 to .35).

Reliability Estimates

Cronbach's alpha was used to measure the internal consistency of the questionnaire. It was pointed out that instrument items should be related to other items if they measure a single construct (Crocker & Algina 1986). Based on the factor analysis conducted above, each factor extracted was treated as a subscale of the EQAO Student Questionnaire, and the internal consistency estimate of each subscale is presented in Table 7.

Cronbach alpha values obtained at the subscale level were relatively low as shown in the above table, ranging from .34 for subscale 3 to .64 for subscale 5. Some of the values did not reach the acceptable level, indicating that these subscales can not be interpreted as internally consistent (e.g., $\alpha > .5$, Kraayenoord & Schneider, 1999). For example, subscale 3 e-literacy activities and subscale 6 literacy hours had Cronbach's alpha values of .34. This phenomenon can be partially explained by the lack of variability in the above subscales, which may be caused by the limit of the original questionnaire variables (e.g., 19 out of the 22 variables were collected on a dichotomous scale).

Table 7

Reliability Estimates

Subscales	Number of items	Items used	Reliability estimates
Subscale 1 Newspaper and magazine literacy activities	4	5, 15, 14, 7	.58
Subscale 2 Non-fiction literacy activities	5	1, 6, 10, 18, 21	.47
Subscale 3 e-literacy activities	3	3, 16, 23	.34
Subscale 4 Song and poetry literacy activities	2	19, 9	.63
Subscale 5 Literature literacy activities	4	8, 20, 13, 12	.64
Subscale 6 Literacy hours	2	11, 22	.34
Subscale 7 Letter reading and writing	2	4, 17	.54

However, owing to the distinctive features that each subscale carried to represent different constructs of students' after-school literacy activities in this research context, all of the seven factors were retained in the subsequent multiple regression analyses to explore the relationship between these constructs of ESL/ELD students' after-school literacy activities and their OSSLT performance.

Multiple Regression

Regression factor scores from the factor analysis were used in the regression analysis to predict which factor(s) contributed to literacy performance. Regression factor scores were used because this procedure automatically converted the values from the related variables into the same metric by using the standard scores (M=0, SD=1). The missing values were handled by excluding cases listwise. An inspection of the data

showed that the data were quite complete. Hence the use of listwise deletion would not result in a large decrease in the sample while maintaining a consistent sample size across all the variables. Also, the dichotomous nature of students' responses to the questionnaire made the common alternative of dealing with missing data (e.g., replacing with the mean) a less practical method of practice. The results of the multiple regression for both reading (M=114.41, SD=33.49) and writing results (M= 104.29, SD= 38.83) are provided in Table 8 on the next page.

When the seven factors were used as the independent variables in the regression analysis, the results were similar for predicting both the reading and the writing scores. All of the seven factors were significant predictors of the OSSLT achievement. Five factors were positively predictive and two were negatively predictive. To be specific, e-literacy activities, literature literacy activities, non-fiction literacy activities, newspaper and magazine literacy activities, literacy hours were positive predictors of ESL/ELD students' performance. Letter reading and writing, song and poetry literacy activities were two negative predictors.

Table 8

Multiple Regression of After-school Literacy Activities and OSSLT Performance

		β	t	Sig.	R^2
Total reading score (Total=200, M=114.41, SD=33.49)					.17
Factor 3	E-literacy activities	.28	19.20	.00	
Factor 5	Literature literacy activities	.17	11.87	.00	
Factor 2	Non-fiction literacy activities	.13	8.76	.00	
Factor 1	Newspaper and magazine literacy activities	.12	8.17	.00	
Factor 6	Literacy hours	.10	7.15	.00	
Factor 7	Letter reading and writing	-.14	6.00	.00	
Factor 4	Song and poetry literacy activities	-.07	-5.05	.00	
Total writing score (Total=180, M=104.29, SD= 38.83)					.10
Factor 3	E-literacy activities	.21	13.93	.00	
Factor 5	Literature literacy activities	.17	11.30	.00	
Factor 2	Non-fiction literacy activities	.09	5.80	.00	
Factor 6	Literacy hours	.07	4.57	.00	
Factor 1	Newspaper and magazine literacy activities	.06	3.87	.00	
Factor 7	Letter reading and writing	-.1	-6.44	.00	
Factor 4	Song and poetry literacy activities	-.06	-4.08	.00	

Note. $* p<.01$

As shown in Table 8, the five positive predictors made their relative contributions in the following order. E-literacy activities was the best predictor of students' performance (for reading β =.28, $p < .01$; for writing β =.21, $p<.01$). To illustrate, the

results suggested that for reading .28*SD (.28*33.49) increase in reading score would occur with a one SD (1) increase in e-literacy activities. The second best predictor was literature literacy activities (for reading $\beta = .17$, $p < .01$; for writing, $\beta = .17$, $p < .01$), which indicated that with the presence of e-literacy activities, literature literacy activities would contribute the greatest amount of unique relevant variance to the regression equation. A one SD increase in literature literacy activities would result in an increase of .17*SD increase in the reading and writing scores, with the contribution of e-literacy activities in the regression equation. Subsequently, the standardized coefficients decreased in the order of non-fiction literacy activities, newspaper and magazine literacy activities, and literacy hours. To be more explicitly, each of the above beta weight denote a unique contribution of the variable while partialing out all the previously entered independent variable(s) in the equation.

In addition, it was also found that this set of factors better predicted the reading scores than the writing scores. Together, 17% of the variance in the reading performance was accounted for by the predictors in the current multiple regression, while only 10% of the variance in the writing performance was explained by the same predictors. The results demonstrated that a substantial variance in ESL/ELD students' literacy performance, especially in writing performance, remained unexplained in the present models.

Home Language Factor

ESL/ELD students' home language backgrounds were examined to understand if their home language patterns would influence their performance on the OSSLT. Investigations of students' responses to the variables of their first language and language(s) spoken at home were conducted.

As reported in Table 5, 6.82% of the students spoke English as their first language and the rest spoke English as their second language. An Independent-Samples T-Test was then completed to examine if this English as first language variable had any influence on literacy performance. The results (Table 9) showed that the two groups' performance differentiated significantly by this variable in both reading and writing scores (p<.01 for reading; p<.01 for writing). The ESL group outperformed the ELD group by 15.41 points in reading, and 17.10 points in writing.

Table 9

Descriptive Statistics and Independent Samples T-Test of Language Variable

	First Language	N	Mean	S D	Sig.
Reading	English	285	100.47	37.32	.01
	Not English	3893	115.88	32.83	
Writing	English	285	88.81	42.62	.01
	Not English	3893	105.91	38.11	

The results of the examination of home language spoken are presented in Table 10. The descriptive statistics showed that the students who only or mostly spoke English at home (English-at-home group) obtained the lowest reading and writing scores. In contrast, the students who only or mostly spoke another language at home (another-language group) performed the best in both reading and writing. The mean score differences of these two groups were 9.61 points for reading, and 14.62 points for writing. The group who spoke English as often as another language at home (combination group) scored in the middle of the other two groups. When comparing the literacy performance of the students

categorized by this home language variable, the difference in writing was still larger than the difference in reading as shown in Table 10.

Table 10

Descriptive Analysis of Home Language Spoken

	Group Categories	N	Mean	S D	Std. Error
Reading	English-at-home group	249	108.67	35.02	2.22
	Combination group	1435	110.13	33.09	.87
	Another-language group	2481	118.28	32.94	.66
	Total	4165	114.90	33.37	.52
Writing	English-at-home group	249	93.84	41.38	2.62
	Combination group	1435	100.48	38.91	1.03
	Another-language group	2481	108.46	37.65	.77
	Total	4165	104.83	38.59	.60

A one way ANOVA (see appendix V) was conducted to determine if significant test performance differences existed among the three groups. The between-group differences for reading was significant ($F (2, 4162) =32.21$, $p<.01$). The between-group differences for writing was also significant ($F (2, 4162) =30.63$, $p<.01$). This demonstrated that literacy performance on the OSSLT differed significantly based on the language(s) spoken at home.

Post hoc tests were completed in order to further determine which groups had significant mean score differences (also see Appendix V). For the reading scores, all the between groups differences were statistically different ($p<.01$), except that the difference

between the English-at-home group and the combination group was not. For the writing

scores, all of the comparisons were significantly different ($p<.01$). This indicated that the

patterns of mean score differences in reading and writing differed across the home

language(s) spoken.

Summary

The examination of the students' questionnaires revealed two aspects. First, five

factors of after-school literacy activities were positive predictors of the OSSLT

performance, including e-literacy activities, literature literacy activities, non-fiction

literacy activities, newspaper and magazine literacy activities, and literacy hours. While

relatively fewer students engaged in letter reading and writing, song and poetry literacy

activities, these factors were negative predictors of literacy performance in the presence

of other factors. Second, the investigation of home language factors further separated ESL

students from ELD students. The sub-groups specified by the two home language

variables demonstrated significant performance difference in both reading and writing.

School-level Indicators and Aggregated School Performance

For the purpose of regressing ESL/ELD students' reading scores and writing scores

on the school-level SES indicators, the two literacy scores were aggregated according to

school identity numbers, resulting in an average reading score of 113.18 (SD=14.40), and

an average writing score of 103.55 (SD=16.27). Also, as noted in the Methods Chapter,

an overall index of the parents' annual income and an overall index of the parental

education level for each school were obtained by calculating the relevant variables in the

original data, resulting in an average income index of 3.14 (SD=.55), and an average

education index of 3.20 (SD=.16). The relationship between both the dependent variables (reading and writing scores) and the independent variables (education index and income index) was analyzed at school level using multiple regression.

First, the data was checked for collinearity because collinearity has adverse effects on regression analysis. Collinearity becomes a major concern when a correlation is above .90 (Tabachnick & Fidell, 2001). To investigate if collinearity existed in this dataset, two values were reported here: correlations and tolerance values. Even though these two values are both based on correlation coefficients, I chose to report both to offer a better understanding of the concern of collinearity in this data. The results indicated that the two SES indicators, the education and the income indices, were correlated, but collinearity was not much a threat. More specifically, the correlation between the education index and the income index was .71. The tolerance value, another collinearity diagnostics index, was .50 (tolerance formula was: $1-R^2$ in the case of two independent variables, therefore, it was $1-.71^2$). Small tolerance values (close to zero) are regarded as symptomatic of collinearity. In this case, it was moderately far away from zero. Therefore, the two SES indicators were both used as independent variables in the regression analysis to predict the aggregated school-level OSSLT performance.

The results (see table 11) indicated that the education index, a variable with low variability (SD=.16), was a significant positive independent predictor of school reading scores ($\beta = .45$, p<.01). In other words, when standardized, a one SD unit increase in the education index was associated with a .45*SD increase in the average school ESL/ELD students' OSSLT reading score. Further, this variable explained 20% of the total variance in the average school OSSLT reading scores for ESL/ELD students.

Table 11

Multiple Regression: Test Performance and School Level SES Indicators

		β	T	Sig.	R^2
1	Education index	.45	5.32	.00	.20
2	Education index	.30	3.30	.00	.09

Note. 1 Dependent Variable: Total Reading Score
2 Dependent Variable: Total Writing Score

As with the total writing scores, the education index also appeared to be a

significant predictor of the school writing scores ($\beta = .30$, p<.01). This indicated that the

education index of a school positively predicted the school writing performance of

ESL/ELD students on the OSSLT. The variable alone accounted for 9% of the total

variance in the writing score. The income index, however, was an insignificant predictor

in both reading and writing in the presence of the education index. Thus it was not

included in Table 11. The reason of the insignificance may be partly explained by the way

stepwise multiple regression was carried out, that is, it first selected the best predictor (the

education index). In the second step of selection, by partialing out all variables already

included in the previous regression equation (the education index), it calculated the

unique relevant variance that the next variable made (the income index). With relatively

high correlation between the income index and the education index, this result was not

surprising. The implications of these findings are presented in the next chapter.

Comparison of Performance in Reading Test Formats

Descriptive Analysis

For the purpose of comparing reading performance in the three formats between ESL/ELD students and non-ESL/ELD students, the descriptive statistics were run using the raw test data. The results are reported in Table 12.

Table 12

Descriptive Statistics of Reading Test Formats

ESL/ELD students (n= 4,311)	Mean /Total (Correct %)	SD	Skewness	Kurtosis
Multiple-choice (MC)	47.87/80 (59.75%)	12.26	-.14	-.40
Constructed Responses (CR)	40.77/70 (58.24%)	13.80	-.49	-.34
CR with Explanations (CRE)	25.77/50 (51.54%)	9.85	-.33	-.48
Non-ESL/ELD students (n=5,003)				
Multiple-choice (MC)	59.30/80 (74.13%)	12.46	-.84	.26
Constructed Responses (CR)	50.89/70 (72.70%)	12.54	-1.11	.99
CR with Explanations (CRE)	32.61/50 (65.22%)	9.05	-.86	.38

The results showed that both the ESL/ELD students and non-ESL/ELD students obtained the highest mean score on the MC questions. The total percent of the correct answers was 59.75% for ESL/ELD students, and 74.13% for non-ESL/ELD students. Both groups achieved less percent of the correct scores in the CR questions, 58.24% for ESL/ELD students, and 72.70% for non-ESL/ELD students. As to the CRE questions, both groups achieved the lowest percent of the correct scores, 51.54% for ESL/ELD

students, and 65.22% for non-ESL/ELD students. The differences on these three formats between the ESL/ELD students and the non-ESL/ELD students were 14.38% for MC questions, 14.46% for CR questions, and 13.68% for CRE questions, suggesting that the performance differences were relatively close on the three reading formats between the two groups.

Furthermore, as indicated by the standard deviations, ESL/ELD students were more spread out on the CR and CRE test formats as compared with non-ESL/ELD students. The standard deviations of ESL/ELD students on these two formats were 13.80 and 9.85, while non-ESL/ELD students' standard deviations were 12.54 and 9.05. An examination of skewness of the two groups in this study demonstrated that non-ESL/ELD students' performance had larger negative skewness value in all three formats, which indicated that non-ESL/ELD students' scores were more clustered together at the higher end of the score distribution. Further, by examining the kurtosis value, it was found that ESL students had negative kurtosis in all three formats (a flat distribution), as opposed to the positive kurtosis obtained by non-ESL students (a peaked distribution).

Discriminant Analysis

Discriminant analysis was conducted to examine which test format had a better discriminating effect between ESL/ELD students and non-ESL/ELD students. The results showed that only MC questions obtained significant standardized canonical discriminant function coefficients ($\beta = 1$, p<.01) (see Table 13). The other two formats did not have significant discriminant coefficients and were thus excluded from Table 13.

Table 13

Discriminant Functions

	Standardized Canonical Discriminant Function Coefficients	Wilks' Lambda	Eigenvalue	Canonical Correlation	R^2	Sig.
MC Questions	1.00	.83	.21	.42	.17	.00

Note. $* p < .01$

In the case of discriminant analysis, in order to have better discriminant functions, eigenvalues should be large. In other words, the ratio of the between-groups sum of squares to the within-groups sum of squares should be a maximum. In the current output, the eigenvalue was .21, which was relatively small. Using the square of the canonical correlation (0.42), only 17.64% of the variance was associated with the differences between groups. Furthermore, Wilk's Lambda was 0.83, indicating that the MC format could provide significant distinction between ESL/ELD students and non-ESL/ELD students, but a large proportion of the total variance was attributable to the differences within groups. In sum, the low eigenvalue coupled with the relatively high Wilk's Lambda indicated that while significant, MC format did not strongly differentiate the status of ESL/ELD and non-ESL/ELD students.

Follow-up correlation analysis revealed that the correlations among the three test formats MC, CR, and CRE were high (.84 between MC and CR; .80 between MC and CRE; .88 between CR and CRE). The high correlations amongst the three item formats indicated that even though students were tested using three different formats, the scores on each of the formats were highly correlated. With MC having the most variance in the

three test formats in discriminant analysis, there left not much unique variance to be shared by the other two formats. In other words, in this study, the difference amongst test formats did not account for much of the variance between ESL/ELD students and non-ESL/ELD students' OSSLT performance.

Table 14

Classification Results

	ESL/ELD status	Predicted Group Membership		Total
		ESL/ELD	Non-ESL/ELD	
Count	ESL/ELD	2785	1526	4,311
	Non-ESL/ELD	1368	3635	5,003
%	ESL/ELD	64.60	35.40	100.0
	Non-ESL/ELD	27.34	72.66	100.0

Note. 68.63% of original grouped cases correctly classified.

Given these results, it was not surprising that the classification results (see Table 14) demonstrated that with the current discriminant function, test format did not prove to be a very good discriminator in separating ESL/ELD students' performance and that of non-ESL/ELD students on the OSSLT. The rows in this table are the observed categories of the dependent variable (ESL/ELD status), and the columns are the predicted categories of the ESL/ELD status. The percentage of cases on the diagonal is the percentage of correct classification. Therefore, only 64.60% ESL/ELD students (2785 out of 4,311) were correctly classified into their correct group, which was 14.60% better than the chance level. Only 72.66% of non-ESL/ELD students (3635 out of 5,003) were correctly classified into their correct group, which was 22.66% better than the chance level.

All together 68.63% of the original grouped cases were classified into the correct group (18.63% better than the chance level), which meant that if using the current discriminant function of test formats, about 35.40% of ESL/ELD students (1526) would be mistakenly grouped as non-ESL/ELD students. About 27.34% non-ESL/ELD students (1368) would be mistakenly classified as ESL/ELD students. Together, these results suggested that test formats could be used to help predict ESL/ELD membership, however, only as a weak predictor.

Summary

The results provided tentative conclusions with respect to the issue of how ESL/ELD students performed on the three reading test formats as compared to their non-ESL/ELD counterparts. The raw score descriptive analysis informed us that in terms of the difficulty levels of the three test formats, the patterns of both groups were the same except that ESL/ELD students performed substantially lower than non-ESL/ELD students. Both groups achieved the highest percent of correct answers on MC questions, and the lowest on CRE questions. In addition, discriminant analysis showed that MC questions obtained the only significant discriminant coefficient. However, given the lack of a strong association between test formats and students' ESL/ELD status, test format did not provide a systematic separation of ESL/ELD students and non-ESL/ELD students.

Chapter Five: Discussion and Conclusion

Overview

The study was conducted to investigate the factors associated with ESL/ELD students' performance on the OSSLT. Previous research has indicated that research investigating how different groups of test-takers performed in tests are a long-standing interest among language testers (Bachman, 2000). This line of research usually provides a context for examining and understanding students' test scores and focuses on their academic background, native language, culture, and ethnicity (Kunnan, 1998). With regard to ESL/ELD students, socio-linguistic history, migration history, past schooling experiences in their native country, motivation, and attitudes towards their new culture are additional aspects that should be included in determining the magnitude and causes of low achievement (Klesmer, 1994). The present study expanded on previous work by examining more factors associated with ESL/ELD students' performance on the OSSLT. The factors included ESL/ELD students' after-school literacy activities, home language patterns, average school-level socio-economic status, as well as test format differences.

Broadly, the results of the study supported the following conclusions. The factors of after-school literacy activities varied in their degrees in predicting ESL/ELD students' OSSLT performance. ESL/ELD students' school performance of the OSSLT could be positively predicted by the average school parental educational index. Further, ESL/ELD students displayed substantial performance gaps compared to non-ESL/ELD students, and test formats provided significant yet weak separation of ESL/ELD students and non-ESL/ELD students. Each of these conclusions is presented below

accompanied by a discussion. Then drawing from the discussion, the implications as well as limitations are presented. The last section is the conclusion.

After-school Literacy Activities and Home Language Patterns

First of all, the results showed that compared with after-school writing activities, ESL/ELD students were more likely to be involved in after-school reading activities. Most students indicated that they spent more time on after-school reading. One reason may be that the majority of the households possessed English-language reading materials at home and they were readily available (see Table 3). Further, writing is viewed as a threefold act that consists of text analytic, composing process, and social construction (Cumming, 1998). Compared with reading as a receptive skill, writing in a second language is a productive skill, which is more cognitively demanding, and requires students' deeper processing of information (McGilly, 1994). In other words, writing is perceived to involve more of the students' initiative in learning and is thus less frequently practiced at home. Some further analyses conducted in this study also demonstrate that, though reading and writing are two closely interrelated aspects of literacy activities, they display quite different features in performance. For example, in terms of the influence of home language patterns, writing seems to be more easily influenced by this variable (see the post-hoc test in Appendix V), and ESL/ELD students showed a larger performance discrepancy as compared to non-ESL/ELD students. Also, the findings of the school-level variables reveal that less variance was explained in writing (9%) than in reading (20%) (see Table 9), which suggests that more elements in writing remained unexplained in the present models. Further, the lower association between writing and family status suggest that writing may be more susceptible to teaching than reading.

The study also reveals that e-literacy activities is an important element of ESL/ELD students' after-school literacy activities, and it has the best predictive effect on these students' literacy achievement. The high percentages of students' engagement in computer-related literacy activities (see Table 3 and Table 4) indicate that e-literacy activities is changing the means or media of how ESL/ELD students read and what they read. Given the importance of computer use on the prediction of literacy performance in this study, further examination needs to be carried out to investigate specifically the effects of students' social communication via the internet in English. Studies focusing on the writing of emails and its effects on writing test performance partially support the contribution of e-literacy activities on the students' performance in this study (see Lee, 1997; Pennington, 2003; Yau, 1994). Lee (1997) pointed out that emails provided students with communicative opportunities for collaborative learning and created a relatively non-threatening atmosphere for students to express, negotiate, and interpret meaning within a meaningful context, which in turn contributed positively to their writing achievement in tests. Pennington (2003) claimed that student attitude was positively promoted by students becoming involved in computer-related literacy activities. Both the quantity and quality of their literacy activity via internet might be positively affected, e.g., writing or reading for extended period of time, forming better revising behavior.

In addition to the significant contribution that e-literacy activities made, it was found that literature literacy activities, non-fiction literacy activities, newspaper and magazine literacy activities, and literacy hours were also positive predictors of the students' performance on the OSSLT. Their contributions were relatively close. Based on these results, tentative conclusion are made that the exposure to and the length of literacy activities outside school are important in relation to literacy test performance. Literature

offers some evidence for this conclusion. For example, Krashen (1985) put forward the view that extensive comprehensible input was crucial to language acquisition. Some empirical studies show that the extent of knowledge and exposure to the English literacy materials are key factors in language development (see Montigny et al., 1991). Caldwell and Gaine (2000) found that the best predictor of reading achievement was determined by the amount of time students spent reading on their own. More specifically, they concluded that one third or more of students' vocabulary growth could be accounted for by after-school independent reading, as it helped to provide students with a wide range of background knowledge, and in turn enhanced their reading achievement in tests. Similarly, extensive outside reading was considered to be helpful in improving the level of proficiency in a global sense, reinforcing the acquisition of grammar, vocabulary, and rhetorical structure, as well as increasing the general knowledge base (Gradman & Hannania, 1991). Therefore, the students who read extensively out of class attained higher levels of proficiency than those with less after-school reading. However, even though the literature abounds with studies emphasizing the importance of the length and exposure to printed materials in English, no literature has been found examining specifically the relative contributions that different literacy materials make to test performance as addressed in the present study. The results, therefore, are open to debate and require further verification.

A look into the text types on the OSSLT, however, provides some alternative explanations for the positive association of the above literacy types with test performance. As can be recalled in the OSSLT component descriptions in Chapter One, there are three reading text types. Information type of reading consists of explanation and opinion. Narrative type of reading consists of stories and dialogues. Graphic type of reading

consists of graphs, schedules, and instruction. A tentative connection between the positive predictors of after-school literacy activities and the text types on the test can be identified. To be more specific, non-fiction literacy activities, and newspaper and magazine literacy activities, which are primarily characterized by information-loaded articles, may well fall into the category of information type of reading on the OSSLT. Literature literacy activities, which is primarily featured with stories and novels, may well fall into the category of narrative type of reading on the OSSLT. Thus the conclusion is when students' after-school reading can be connected to the text types on the test; chances are good that the literacy performance on the OSSLT can be positively predicted.

This alternative explanation of the connection between OSSLT text types and students' after-school reading materials may also account for the two negative predictors in the study. Letter reading and writing, song and poetry literacy activities were reported to be less frequently practiced by most ESL/ELD students, and there is no text types on the OSSLT that require students to demonstrate the abilities as required in these literacy activities. On the other hand, it is also possible that many of the ESL/ELD students do not reach the linguistic threshold or literacy level that can allow them to engage in poetry literacy activities or the like, as these kinds of literacy activities are more culturally embedded and linguistically demanding for ESL/ELD students to practice on their own outside school.

Complex home language patterns notwithstanding, the results of the mean difference tests between ESL students and ELD students (see Table 11), suggest that ELD students, who lack a formal education in either home language or English, are the students who performed worse than ESL students on this literacy test. Though no research has focused particularly on ELD students, the finding, however, is supported by studies

from second language education. For example, Gardner et al. (1996) maintained that years of formal education in the home country was a significant predictor of students' academic success (see also Gradman & Hannania, 1991; Willms, 1997). They argued that well-developed literacy skills from the first language helped the acquisition and development of the second language. This result can also be explained by some theoretical discussion in second language education. For example, Linguistic Interdependence Hypothesis (LIH) (Alderson, 1984), which suggests that reading performance in a second language is largely shared with reading ability in a first language. The Linguistic Threshold Hypothesis (LTH), which means that a level of second language linguistic ability must be obtained in order to achieve efficiency in the development of a second language (Bernhardt & Kamil, 1995). Common Underlying Proficiency (CUP) model (Cummins, 1996), which identifies that there is an underlying cognitive/academic ability that is common across languages. A large mass of higher-order thinking skills, including analysis, synthesis, interpreting, reasoning, generalizing, and transferring may be transferable across languages. The theoretical discussion of the first language influence on the second language development informs that first language abilities provide ESL students with heuristic frameworks to help mediate the second language process (Hudson, 1998). Thus with formal education in the first language, ESL students possess a considerable amount of information about their first language system, skills and learning strategies and a wide range of concepts on which to draw when confronted with new words in a second language. In other words, ESL students are more cognitively advanced in terms of language development, which helps them to transfer their cognitive capacity to the development of their second language. In contrast, ELD students could not resort to those skills or strategies that ESL students possess. They are

cognitively weak in language(s), no matter in English or in other language(s). Therefore, the current study concluded that ELD students, though being a small portion of the student population (6.82% of 4311 ESL/ELD population), were the group that was one of the most at-risk on this high-stake literacy test.

However, inconsistent results arose with regard to ESL/ELD students' home language patterns. There were more students who claimed that English was their first language than English-at-home group (see Table 3) even though both numbers were relatively small. The reasons for this discrepancy (0.84%) remain unclear. It may be due to measurement error or random effect. It can be reasonably assumed that English as first language students would speak English at home in this English-speaking country. If this premise is true, the numbers of English-at-home group should be larger than the numbers of students whose first language was English. Certainly, the real situations of ESL/ELD students' home language patterns could be very complex and not easy to identify. For example, there are possibilities that certain English as second language students belong to English-at-home group due to their family circumstances (e.g., one parent only speaks English), or family influence (e.g., even though English is the family's second language, still only or mostly English is spoken at home).

The most puzzling finding is the difference found with two ESL student groups. Those who mainly speak another language at home outperformed the students who speak a combination of another language and English (see Table 10). It is reasonably assumed that within this context, this "another language" is most likely to be the students' first language learned. By reviewing the literature, it is found that much focus is on the L1 influence on L2, e.g., Linguistic Interdependence Hypothesis, Linguistic Threshold Hypothesis and the Common Underlying Proficiency model as discussed above. However,

no further classifications have been made within ESL student groups. Therefore, this finding definitely needs to be cross-validated with other ESL population in order to know more clearly how home language would influence on ESL students' English literacy development, who vary to a large extent in their home language patterns. For instance, homestay international ESL students, though speaking English as second language, would possibly check the box of "another language as often as English". It is also quite reasonable to attribute the result to this complex and varying situations of ESL students.

Some thoughts are offered from the discussion of the two aspects of language proficiency in second language education research: conversational English proficiency and academic English proficiency. Cummins (1981) refers to conversational English proficiency as Basic Interpersonal Communicative Skills (BICS), which is usually context-embedded and cognitively undemanding. Academic English proficiency is referred to as Cognitive/Academic Language Proficiency (CALP), which is context-reduced and cognitively demanding. Obviously, students are required to exhibit their academic English on the OSSLT, which may not necessarily be equal to or transferable from their daily conversational English. Cummins (1989) pointed out that it is those students who are in the transition from BICS to CALP that need to be called into special attention, as their "sound good" oral English may quite possibly disguise their lack of development in higher levels of reasoning and integrative language skills required in tests. Considering that ESL students who wrote the OSSLT were still in their first few years of stay in Canada, this explanation may be tenable but requires empirical investigation. In sum, the finding regarding the home language variable and literacy performance is complicated. L1 is supporting the development of L2, but how L1 is supporting L2 and the extent of the support or impact remain to be probed further.

School-level Socio-economic Variables

As to the school-level socio-economic variables, the results suggested that a high parental education level had a positive association with the average school performance (see Table 11). This finding is in line with previous studies. Mothers' and fathers' education had strong influences on their children's educational outcomes (Finnie & Meng, 2003). Students from more educational advantaged family backgrounds are more likely to have access to quality education, and greater financial and cultural capital to support their educational activities (Willms, 1997). Bianchi and Robinson (1997) coined the term "social capital" to mean that social interaction with parents could create resources that would enhance the child's realization of his or her potential for achievement and later academic success. Parental resources also place children in certain types of communities and educational environments where social interactions could provide further social capital to facilitate "good" student outcomes (p. 333). Wendling and Cohen's (1980) reported that an increase of one year in the median years of schooling of the parent population was associated with a change in students' test score of 1.24 points. Furthermore, their findings revealed that approximately one half of the mean reading score and the mean math score was attributable to this factor.

Research also has shown that parents with higher educational level are accustomed to learning about, keeping track of, and taking advantage of opportunities in the school system. They tend to have the advantage of knowing the importance of providing support to the school (Hickman et al., 1995). College-educated mothers are found to devote significantly more time to children than mothers with only a high school education or less (Bianchi & Robinson, 1997). On the other hand, parents with lower educational levels are

linked to less effective parent-school interactions and increased risk for student academic failure (Griffith, 1996, 1997). In addition, Baker and Stevenson (1986) found that mothers with higher levels of education were more likely to implement strategies for managing their children's progress in schools and were better informed about their children's performance, which in turn helped promote their children's achievement in tests (see also Fehrmann, 1987; Griffith, 1996, 1997; Lytton & Pyryt, 1998).

It is worth noting that given the fact that the education index was correlated to the income index ($r = .71$), the insignificance of the income index in this context did not mean that its influence on students' test performance could be neglected. The examination of the association between the income index alone and students' test performance was significant (the analysis was run but was not reported here). Evidence in literature can be found in Wendling and Cohen's (1980) study that parental poverty had a strong association with lower test scores, such that test scores changed by approximately .25 points with a one percent change in the poverty population (see also Willms, 1997).

In considering the findings from both student-level and school-level variables together, the major findings are consistent with the EQAO 2003 report of the Provincial Results. In that report, the following variables showed the strongest relationship with students' achievement: "access to a home computer, average household income, average household educational attainment, total school enrolment, average stakeholder rating on school safety, average student rating on school safety, and average respondent rating on leadership measures" (p.21-22). In the current study, e-literacy activities showed the strongest relationship with ESL/ELD students' individual literacy achievement. Average parental education index showed the strongest relationship with the aggregated ESL/ELD students' literacy performance. In addition, the students who reported to engage in

e-literacy activities represented a group of students who had relatively higher socio-economic status. Therefore, e-literacy activities, denoting a proxy variable of SES, evidentially supported the close relationship between ESL/ELD students' test performance and their SES background. Social or economic capitals offered to students greatly influence their school success. The findings also corroborated with the view that examining and understanding students' achievement scores should be placed in a broader context, encompassing necessary contextual features, such as students' accessibility of computers and resource materials, and the community/school socio-economic status (EQUIP: The framework, 2004).

<div align="center">Comparison of Test-format Differences</div>

The results of the third research question showed that ESL/ELD students performed worse in all three test formats in the reading section as compared to their non-ESL/ELD counterparts. However, the performance patterns were more or less the same. Both groups achieved a higher percentage of correct answers in MC questions, lower in CR questions, and the lowest in CRE questions.

The findings are partially supported by the literature, that is, MC questions are generally easier to accomplish as compared to CR questions or CRE questions (Fitzgerald, 1978; Wolf, 1993). Wolf's (1993) study also showed that students obtained higher achievement scores on MC questions than CR questions (see also Shohamy, 1984), due to the fact that MC required "comprehension and selection", while CR required "comprehension and production" (p.481). Further, MC questions were usually regarded to be conducive to "test-wiseness", a series of general strategies to efficient test-taking (Bachman, 1990), e.g., ruling out as many alternatives as possible and then guessing

among the ones remaining. Such strategies may have resulted in higher scores in MC compared with in the formats of CR and CRE in this study. Further evidence was provided in Cheng and Gao's (2002) experiment. They found that in doing MC questions in reading comprehension, even in the case of not providing the associated reading passages, second language students achieved scores above the chance level. Although most of the studies on test format differences are conducted with second language students, the same performance pattern in relation to MC and CR or CRE formats applied to both ESL/ELD and non-ESL/ELD students in this study. In addition, the finding that ESL/ELD students' performance were more spread out than non-ESL/ELD students in this study demonstrate that ESL/ELD students, though all in a language developing process, varied considerably in their literacy achievement process. Therefore, attention should be paid to examine individual differences among the group of ESL/ELD students.

The discrimination effect regarding test formats suggest that there is a lack of association between test formats and test performance. Also, the unsatisfactory classification results based on the discrimination functions (see Table 14) provide further evidence demonstrating the weak discriminating power that test formats had in this context. Given the current discrimination function based on test formats, around one third of the students could not be assigned to their correct group. By considering the results of descriptive analysis and discriminant analysis together, the results suggest that there are large performance gaps between ESL/ELD students and non-ESL/ELD students, but these gaps can not be attributed much to the test format difference for this study. Cheng, Klinger, & Zheng (under review) conducted a two-year cross validation study of the OSSLT data. Their results showed that the discrimination effects regarding test formats were not consistent over the two years. For the February 2002 data, CR questions best

separated the two groups ($\beta = .42$, p< .01). MC questions had a discriminant coefficient

of .34 (p< .01), and CRE question had the lowest discriminant coefficient of .30 (p< .01).

For the 2003 data[6], only MC items best separated the two groups ($\beta = 1$, p< .01), which

is the same as this study. Also, they found that the performance differences were smaller

in 2003 compared with 2002. The possible explanation they offered was that February

2002 was the first test administration, and October 2003 data was the third administration.

Thus the smaller performance differences in part reflected the progress that ESL/ELD

students made, or the extent that these students were coached for the test. Also, the

smaller performance differences in 2003 might lead to the diminishing of the

discrimination effects of the other two test formats, leaving only the MC format as a

significant discriminator. Therefore, test format did not provide a systematic separation

between ESL/ELD and non-ESL/ELD students. There seemed to be a lack of strong

association between test performance and test formats in this study, and the most difficult

constructs did not necessarily coincide with the best discriminator.

Even though a systematic analysis of the actual OSSLT items is necessary to gain a

deeper understanding of this result, possible explanations are offered here: the CR or CRE

questions on the OSSLT might not substantively require students to employ deeper

cognitive levels, e.g., synthesis level, or to apply sophisticated background knowledge

(which would place ESL/ELD students at a disadvantage) to correctly answer the CR or

CRE questions. Thus whether the actual test items of CR and CRE questions on the

OSSLT supports the following arguments from the literature reviewed above deserve

further investigation. These arguments are: (1) constructed-response formats are more

---

[6] The 2003 data used in Cheng, Klinger & Zheng's study were the same data that I used in my study.

advanced in assessing dynamic cognitive processes, as they are capable of asking students to employ not only students' knowledge-level, but also synthesis-level cognition (Bloom et al., 1970); (2) constructed-response formats replicate more faithfully the tasks test-takers face in academic and work settings (Bennett et al., 1990).

To summarize, the results of this study confirmed that ESL/ELD students displayed substantial performance discrepancies as compared to non-ESL/ELD students. These discrepancies, however, are close on each format, indicating that group performance difference exists due to a real difference in the ability being tested or some other unknown factors, rather than the effects of test method. In order to know if test formats make a difference, it is important to combine the analysis of test performance and the examination of how the test formats are constructed and how the questions are answered by different groups of students. Test formats alone do not justify the cognitive levels that they attempt to tap, or the techniques that need to be employed in answering the questions.

<div align="center">Implications and Limitations of the Present Study</div>

From the preceding discussion, there are some implications and limitations of this study, in at least the following areas.

Implications for Research

This study provided some necessary information for addressing the confounding variables that were associated with ESL/ELD students' performance on the OSSLT. The study also demonstrated how to evaluate the impacts of these types of variables. However,

the link between these characteristics is not yet well understood. The following sections outline the implications of the study for research purposes and for practical purposes.

As can be seen, the definition or scope of literacy is rapidly changing (Lankshear & Knobel, 2003). Many of the new and changing social practices would involve new and changing ways of producing, distributing, exchanging and receiving texts by electronic means (Lankshear & Knobel, 2003). The results of the significant effect of students' e-literacy activities should evoke some further investigations into the issues associated with the massive growth of electronic information and communications technologies and their increasing roles in students' literacy development. The fact that today's society is surrounded with computer use in daily life, especially in students' lives implies that more research should look into the process of reading information through the internet, writing e-mails, or exchanging messages through text messenger (e.g., MSN messenger, Yahoo messenger). In addition, given the fact that e-literacy activities is conducted over different medium, unlike the medium employed on the OSSLT, the exact nature of the impact of e-literacy activities on literacy performance calls for more studies to be conducted to uncover whether the impact pattern is unique to these ESL/ELD students.

Further research regarding differentiating ELD students and ESL students should be conducted, especially in the area of capturing their unique literacy characteristics and literacy performance, in order to provide more concrete and systematic help to these students, especially to ELD students. It should also be noted that the status of ESL/ ELD students is a fluid concept. It changes with the change of the students' length of stay in Canada, their family situations, and their English literacy development. Research needs to confirm empirically how first language literacy impacts on the development of a second

language literacy during different stages, especially in a country like Canada, in which there are a variety of students with various language backgrounds.

In addition, a comparison of SES effects on ESL/ELD students and on non-ESL/ELD students would be worthwhile to investigate. Some studies have shown that certain ESL families may have low socio-economic status, but value highly their children's education. For example, Fehrmann (1987) found that by controlling other background characteristics, minority students have parents with higher levels of involvement in children's learning when they are being compared with their majority counterparts from the same SES background. Also, in the present study, ESL students were treated as a homogenous group, which may quite possibly disguise the difference within ESL students (see also Finnie & Meng, 2003). It is, therefore, worthwhile to compare ESL students' performance from other jurisdictions, such as previous schooling experiences, previous exposure to the English language, and SES background.

Implications for Practice

In terms of providing help to ESL/ELD students, it is essential for teachers to know the following about their students. What are ESL/ELD students' literacy characteristics? How would first language impact on the development of a second language? And what are the specific performance patterns on different test formats that ESL/ELD students would display so that teachers could offer help? For parents, even though SES is a type of factor that is not easily manipulated, it is important for parents to realize that the amount of support they offer is in fact countable (Fehrmann, 1987). For example, with the knowledge of the positive association between students' after-school literacy activities and their literacy performance, it is essential for parents to know that by guiding their

children to engage in literacy activities that are more connected to text types on the OSSLT would help them to perform better (e.g., literature literacy activities, non-fiction literacy activities). Also, helping students to foster good after-school reading habits or engage in more meaningful e-literacy activities would be beneficial to their children's literacy development. It is also vital for parents to realize that the development of a first language is unlikely to be a restraint in the development of a second language. Instead, first language literacy is an asset (Hudson, 1998), and certain cognitive skills are transferable among languages, hence are beneficial in ESL students' English literacy development.

Limitations and Suggestions

It is important to note that there are limitations to the present study. First, the questionnaire used in this study only gave an all overall picture of the ESL/ELD students' after-school literacy factors, and the questions asked are mostly general in nature. Given the relatively low reliability estimates of the questionnaire (see Table 7), the constructs of literacy activities need to be further refined, and questions need to be polished to better represent the constructs of after-school literacy activities. In addition, the questionnaire is designed mostly on dichotomous scales rather than on continuous or ordinal scales (e.g., likert scale), which sets limits to the exploration of the impact of the designed variables.

In the meantime, it should be noted that information about ESL/ELD students' background, such as country of origin, the length of stay in Canada/English-speaking countries, and their motivation and strategies in English literacy development were not collected in the questionnaire. These variables are regarded as important for understanding literacy development of ESL/ELD students as suggested in the literature

(Kunnan, 1998; Klesmer, 1994; Yau, 1994). Most importantly, a broader qualitative investigation would be a good supplement to probe into different ethnic groups or different levels of achievement groups. A better understanding needs to be developed of ESL/ELD students' test performance, literacy characteristics, and their adjustment to the Ontario high school testing system. This understanding will provide educators, teachers, and parents a better and deeper insight into the actual performance and needs of different groups of ESL/ELD students.

The second limitation lies in school-level variables. SES measures, which were applied to schools, not to individuals, paint a broad-brush picture that may quite possibly hide some differences in income or educational attainments among individual ESL/ELD students. Also, if the data had a stratified collection of the individual ESL students' SES status, then a more detailed examination of the relationship between ESL/ELD students' performance on the OSSLT and their family factors could be conducted. More importantly, this SES information could be connected to other test-taker characteristics, including affective or cognitive measures. Also non-ESL/ELD students' test-taker characteristics were not examined in this study to allow a comparison study to be conducted to understand whether the relationships found so far were unique to ESL/ELD students.

The lack of systematic analysis of the OSSLT items is another limitation of this study, especially in comparing ESL/ELD students and non-ESL/ELD students' performance in the reading section. Systematic analysis of the actual test items would not only provide more concrete information about the test and the test-takers, but promote the significance of similar studies using sophisticated testing theories. Also, this study needs

to be replicated with other OSSLT administrations in order to cross validate the findings of this study.

Conclusion

The paucity of literature that explores ESL/ELD students' performance on high-stakes tests suggests the need for further studies. These students are in a transition period of their lives, during which they may pursue higher education or enter the job market. Valid and reliable tests of their literacy capability would serve as an effective incentive to foster more efficient ways of reading and writing in and outside the classroom, and thus enhance the human capital of the nation in the long term (Willms, 1997). Also, as noted, ESL/ELD students make up a substantial and growing portion of the Canadian high school population. It is therefore important to analyze the challenges that ESL/ELD students face and identify the factors and barriers associated with their literacy development in students' literacy development.

The results of this study showed the relationship of ESL/ELD students' test performance and their after-school literacy activities and home language patterns. The findings highlighted the essential role of computer-related literacy activities in ESL/ELD students' literacy performance and the positive association of the school-level parental education index and ESL/ELD students' aggregated literacy performance. The findings also provided information about the performance discrepancy between ESL/ELD students and non-ESL/ELD students on the reading test formats. In better coaching ESL/ELD students to meet this Ontario high school graduation requirement, more efforts should be made to look not only into test formats, but other aspects, such as reading skills, strategies, text types of reading, and writing tasks to help this less-advantaged group.

To summarize, ESL/ELD students bring increasing linguistic and cultural diversity to Ontario high schools. This phenomenon provides cultural enrichment, as well as poses challenges to the society. Schools or school boards need to use the OSSLT test results along with other sources of contextual information to improve classroom instructions that are needed to facilitate ESL/ELD students' literacy development and cultural adjustment. Schools or school boards should also utilize the information provided to implement improvement planning, such as providing appropriate assessment alternatives (e.g., classroom assessment, portfolio assessment) for the purpose of lowering the risk of jeopardizing ESL/ELD students' school success in Ontario. Seeking for valid and fair evaluation of all students' school achievement should be a never-ending endeavor for education-related professionals to pursue so as to not further marginalize those students who are at the margin of the society. As a province that welcomes a large proportion of the immigrants across the nation, prompt and individualized help to students from diverse linguistic and cultural background should become an inseparable ingredient of Ontario educational system.

References

Alderson, J. C. (1984). Reading in a foreign language: A reading problem or a language problem. In C. Alderson, & H. Urquhart (Eds.), Reading in a foreign language (pp. 1-25). New York, NY: Longman.

Alderson, J. C., & Urquhart, A. H. (1985). The effect of students' academic discipline on their performance on ESP reading tests. *Language Testing, 2,* 192-204.

Bachman, L. F., & Palmer, A. S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly, 16*(4), 449-465.

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

Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing. 17*(1) 1-42.

Baker, D. P., & Stevenson, D. L. (1986). Mothers' Strategies for Children's School Achievement: Managing the Transition to High School. *Sociology of Education, 59*(3), 156-166.

Bennett, R. E., Ward, W. C., Rock, D. A., & Lahart, C. (1990). *Toward a framework for constructed-response items* (ETS Research Report No. 90-7). Princeton, NJ. Educational Testing Service.

Bennett, R. E., Rock, D., & Wang, M. (1991). Equivalence of free-response and multiple-choice items. *Journal of Educational Measurement, 28*(1), 77-92.

Bernhardt, E. B., & Kamil, M. L. (1995). Interpreting relationships between L1 and L2 reading: Consolidating the linguistic threshold and the linguistic interdependence hypotheses. *Applied Linguistics, 16*(1), 15-34.

Bianchi, S. M., & Robinson, J. (1997). What did you do today? Children's use of time, family composition, and acquisition of social capital. *Journal of Marriage and the Family, 59*(2), 332-344.

Bloom, B. S., Englehart, M. B., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. New York: McKay.

Caldwell, K., & Gaine, T. (2000): The phantom tollbooth" and how the independent reading of good books improve student's reading performance (ERIC Document Reproduction Service No. ED 449 462).

Canadian Immigration and Citizenship (2002). *Immigration Overview.* Retrieved on November 5, 2005, from

http://www.cic.gc.ca/english/pub/facts2002/immigration/immigration_7.html

Canadian Immigration and Citizenship (2003). *Immigration Overview: Permanent Residents.* Retrieved on November 5, 2005, from

http://www.cic.gc.ca/english/pub/facts2003/permanent/1.html

Canadian Immigration and Citizenship (2003). *Immigration Overview: Permanent and Temporary Residents.* Retrieved on November 5, 2005, from

http://www.cic.gc.ca/english/pub/facts2003/overview/3.html.

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

Cheng, L., & Gao, L. (2002). Passage dependence in standardized reading comprehension: Exploring the College English Test. *Asian Journal of English Language Teaching, 12,* 161-178.

Cheng, L., Klinger, D., & Zheng, Y. (under review). The Challenges of the Ontario

secondary school literacy test for second language students.

Chihara, T., Sakurai, T., & Oller, J. W. Jr., (1989). Background and culture as factors in EFL reading comprehension. *Language Testing, 6,* 143-163.

Collier, V. P. (1989). How long? A synthesis of research on academic achievement in a second language. *TESOL Quarterly, 23*(3), 509-531.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York, NY: Saunders College Publishing.

Cumming, A. (1998). Theoretical perspectives on writing. *Annual Review of Applied Linguistics, 18*, 61-78.

Cummins, J. (1980). The cross-lingua dimensions of language proficiency: Implications for bilingual education and the optimal age issue. *TESOL Quarterly, 14,* 175-187.

Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics, 11*(2), 132-149.

Cummins, J. (1996). *Negotiating identities: Education for empowerment in a diverse society.* Los Angeles, CA: California Association for Bilingual Education.

Diao, J. (2004). *Factors affecting ESL students' academic learning through English.* Master's project. Queen's University. Kingston, ON.

Early, M. (1992). Aspects of becoming an academically successfully ESL student. In B. Burnaby & A. Cumming (Eds.), *Sociopolitical aspects of ESL* (pp. 265-275). Toronto: OISE Press.

Education Quality and Accountability Office. (2002). *Ontario Secondary School Literacy Test, October 2002: Report of Provincial Results*. Retrieved June 14, 2004, from http://www.eqao.com/pdf_e/02/02P026e.pdf.

Education Quality and Accountability Office. (2003). *Ontario Secondary School Literacy Test, October 2002: Report of Provincial Results*. Retrieved June 14, 2004, from http://www.eqao.com/pdf_e/03/03P006e.pdf.

Education Quality and Accountability Office. (2004). *Ontario Secondary School Literacy Test, October 2003: Report of Provincial Results*. Retrieved January 20, 2005, from http://www.eqao.com/pdf_e/04/04P002e.pdf.

Educational Quality Indicators Program (EQUIP): The Framework. (2004). Retrieved June 16 2004, from http://www.eqao.com/categories/equip_e.aspx.

Finnie, R., & Meng, R. (2003). *Minorities, cognitive skills and the incomes of Canadians.* (Rep. No. 196-11F0019). Ottawa, ON: Statistics Canada.

Fisher, T. A. (2000). Predictors of academic achievement among African American adolescents. In S. T. Gregory (Ed.), The academic achievement of minority students: Perspectives, practices, and prescriptions. (pp. 307-334). Lanham, Maryland: University Press of American.

Fitzgerald, T. P. (1978). A cross cultural study of three measures of comprehension at the primary and intermediate levels. *Educational Research, 3*(2), 84-92.

Fehrmann, P. G. (1987). Home influence on school learning: direct and indirect effects of parental involvement on high school grades. *Journal of Educational Research, 80*(6), 330-337.

Gardner, S., Polyzoi, E., & Rampaul, Y. (1996). Individual variables, literacy history, and ESL progress among Kurdish and Bosnian immigrants. *TESL Canada Journal, 14*(1), 1-20.

Garson, G. D. Discriminant function analysis retrieved on May 1, 2005 from http://www2.chass.ncsu.edu/garson/pa765/discrim.htm.

Gay, L. R. (1980). The comparative effects of multiple-choice versus short-answer tests on retention. *Journal of Educational Measurement, 17*(1), 45-50.

Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psychology* (3rd ed.). Boston : Allyn and Bacon.

Gradman, H. L., & Hannania, E. (1991). Language Learning Background Factors and ESL Proficiency. *Modern Language Journal,* 75 (1), 39-51.

Gorsuch, R. L. (1983). *Factor analysis.* Hilldale, NJ: Lawrence Erlbaum Associates.

Graff, H. J. (1994). Literacy, myths and legacies: lessons from the history of literacy. In L. Verhoeven (Eds.), *Functional literacy: Theoretical issues and educational implications* (pp. 37-60). John Benjamins B.V.: Amsterdam.

Griffith, J. (1996). The relation of parental involvement, empowerment, and school characteristics to student test performance: An organization analysis. *Journal of Educational Research, 90*, 33-41.

Griffith, J. (1997). Linkage of school structural and socioenvironmental characteristics to parental satisfaction with public education and students' academic achievement. *Journal of Applied Social Psychology, 27*(2), 156-186.

Hancock, G. R. (1994) Cognitive complexity and the comparability of multiple-choice and constructed-response test formats. *Journal of Experimental Education, 62*(2), 143-157.

Hickman, C. W., Greenwood, G., & Miller, M. D. (1995). High school parent involvement: relationships with achievement, grade level, SES, and gender. *Journal of Research and Development in Education, 28*(3).125-134.

Horn, C. (2003). High-stakes testing and students: stopping or perpetuating a cycle of failure? *Theory to Practice, 41*(1), 30-41.

Hudson, T. (1998). Theoretical perspectives on reading. *Annual Review of Applied Linguistics, 18*, 43-60.

Johnson, T. R., & Krug, K. (1980). Integrative and instrumental motivations: In search of a measure. In J. W. Oller Jr., & K. Perkins (Eds.), *Research in language testing* (pp. 241-249). Rowley: Newbury House.

Klesmer, H. (1994). Assessment and teacher perceptions of ESL student achievement. *English Quarterly, 26*(3), 8-11.

Kobayashi, M. (2002). Method effects on reading comprehension test performance: text organization and response format. *Language Testing, 19*(2), 193-220.

Krashen, S. D. (1985). *The input hypothesis.* London: Longman.

Kraayenoord, C. E. & Schneider, W. E. (1999). Reading achievement, metacognition, reading self-concept and interest: A study of German students in grades 3 and 4. *European Journal of Psychology of Education, XIV,* 305-324.

Kunnan, A. J. (1995). *Test-takers characteristics and test performance: A structural modeling approach.* New York, NY: Cambridge.

Kunnan, A. J. (1998). Approaches to validation in language assessment. In A. J. Kunnan (Ed.), (pp. 1-16). *Validation in language assessment.* Mahwah, NJ: Lawrence Erlbaum.

Lankshear, C. & Knobel, M. (2003). New literacy, changing knowledge and classroom learning. Philadelphia, PA: Open University Press.

Lee, L. (1997). Using internet as an enhancement of C2 teaching and learning. *Foreign Language Annals, 30*(3), 410-425.

Liu, M., Moore, Z., Graham, L., & Lee, S. (2003). A look at the research on computer-based technology use in second language learning: A review of the

literature from 1990-2000. *Journal of Research on Technology in Education, 34*(3), 250-274.

Lukhele, R., Thinssen, D., & Wainer, H. (1994). On the relative value of multiple-choice, constructed-response, and examinee-selected items on two achievement tests. *Journal of Educational Measurement, 31*(3), 234-250.

Lytton, H., & Pyryt, M. C. (1998). Predictors of achievement in basic skills: A Canadian effective schools study. *Canadian Journal of Education, 23*(3), 281-297.

Madaus, G., & Clarke, M. (2001). The adverse impact of high-stakes testing on minority students. In M. Kornhaber & G. Orfield (Eds.), *Raising standards or raising barriers: Inequality and high-stakes testing in public education* (pp. 85-106). New York: Century Foundation.

MaGilly, K. (1994). Classroom lessons: Integrating cognitive theory and classroom practice. Cambridge, MA: MIT Press.

Magiste, E. (1984). Further evidence for the optimal age hypothesis in second language learning. In J. P. Lantolf, & A. Labarca (Eds.), *Research in second language learning: Focus on the classroom* (pp. 51-58). New Jersey: Ablex Publishing Corporation.

McGilly, K. (1994). Classroom Lessons: Integrating cognitive theory. Cambridge, MA: MIT Press.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.). New York: Macmillan, 13-103.

Ministry of Education and Training (1999). *English as second language-literacy development: The Ontario curriculum grades 9 to 12.* Toronto: The Ministry of Education.

Montigny, G., Kelly, K., & Jones, S. (1991). *Adult literacy in Canada: Results of a national study* (Rep. No. 89-525-XPE). Ottawa, On: Statistics Canada.

Mulvenon, S. W., Ganley, B. J. & Fritts-Scott, K. (2001). The impact of socioeconomic status on performance: A case study of Arkansas schools. *Arkansas Educational Research & Policy Studies Journal, 1(1),* 42-61.

Norusis, M. J. (1988). *SPSS-X advanced statistics guide* (2nd ed.). Chicago, IL: SPSS Inc.

Oller, J. W. Jr., & Hinofotis, F. (1980). Two mutually exclusive hypotheses about second language ability: indivisible or partially divisible competence. In J. W. Oller Jr., & K. Perkins (Eds.), *Research in language testing* (pp. 13-23). Rowley: Newbury House.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach.* Hilldale, NJ: Lawrence Erlbaum Associates.

Pennington, M. C. (2003). The Impact of the computer in second language writing. In Kroll, B. (Ed.), *Exploring the dynamics of second language writing.* New York, NY. Cambridge University Press.

Ramirez, A. G. (1994). Literacy acquisition among second-language learners. In B.M. Fredman, R. Weber & A.G. Ramirez (Eds.), *Literacy across language and cultures* (pp. 75-101). Albany, NY: State University of New York Press.

Riley, S., & Lee, J. F. (1996). A comparison of recall and summary protocols as measures of second language reading comprehension. *Language Testing, 13*(2), 173-190.

Roessingh, H. (1999). Adjunct support for high school ESL learners in mainstream English classes: Ensuring success. *TESL Canada Journal, 17*(1), 72-85.

Roessingh, H. & Kover, P. (2002). Working with younger-arriving ESL learners in high school English: Never too late to reclaim potential. *TESL Canada Journal, 19*(2), 1-19.

Shephard, L. (1991). Negative policy for dealing with diversity: When does assessment and diagnosis turn into sorting and segregation? In E. Hiebert (Ed.), *Literacy for a diverse society: Perspectives, practices, and policies* (pp. 279-298). New York: Teachers College Press.

Shohamy, E. (1983). The stability of oral proficiency assessment on the oral interview testing procedure. *Language Learning, 33*, 527-540.

Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing, 1*, 147-170.

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: the need for new research and practice paradigms on the testing of English-language learners. *Educational Researcher, 32*(2), 3-13.

Snow, R. E. (1993). Construct validity and constructed-response tests. In R.E. Bennet & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed-response, performance testing, and portfolio assessment* (pp. 45-60). Hillsdale, NJ: Lawrence Erlbaum Associates.

Statistics Canada (1996). *Reading the Future: A Portrait of Literacy in Canada.* (Rep.No. 89-551-XPE). Ottawa, ON: Statistics Canada & Human Resources Development Canada.

Subrahmanyam, K., Kraut, R. E., Greenfield, P. M., & Gross, E. F. (2000). The impact of home computer use on children's activities and development. *The Future of Children, 10*(2), 123-145.

Tabachinick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics.* (4rd ed.). Boston: Allyn & Bacon.

Tuijnman, A. (2001). Benchmarking adult literacy in North America: An international comparison study. (Rep. No. 89-572-xie). Ottawa, ON: Statistics Canada.

Wainer, H., & Thissen, D. (1993). Combining multiple-choice and constructed-response test scores: Toward a Marxist theory of test construction. *Applied Measurement in Education, 6*(2), 103-118.

Wartella, E. A., & Jennings, N. (2000). Children and computers: new technology—old concerns. *The Future of Children, 10*(2), 31-44.

Watt, D., & Roessingh, H. (1994). ESL dropout: the myth of educational equity. *The Alberta Educational Research, XL*(3), 283-296.

Watt, D., & Roessingh, H. (2001). The dynamics of ESL dropout: Plus ca change! *The Canadian Modern Language Review, 58*(2), 203-222.

Wendling, W., & Cohen, J. (1980). The relationship of education resources to student achievement levels in New York State. (Eric Document Reproduction Service No 202176).

Wolf, D. (1993). A comparison of assessment tasks used to measure FL reading comprehension. *Modern Language Journal, 77,* 473-489.

White, K. (1982). The relation between socio-economic status and academic achievement. *Psychological Bulletin, 91*, 461-481.

Willms, J. D. (1997). Literacy skills of Canadian youth. (Rep. No. 89-552). Ottawa, ON: Statistics Canada.

Yau, M. (1994). Provincial grade 12 writing review: A close look at Toronto's

    performance. Toronto Board of Education, ON. (Eric Document Reproduction

    Service No. ED 377 489).

Zeidner, M. (1986). Are English language aptitude tests biased towards culturally

    different minority groups? Some Israeli findings. *Language Testing, 3*, 80-98.

Zeidner, M. (1987). A comparison of ethnic, sex and age bias in the predictive validity of

    English language aptitude tests: some Israeli data. *Language Testing, 4*, 55-71.

Appendix I: EQAO Student Questionnaire

Ontario Secondary School Literacy Test, October 2003

EQAO
Education
Quality and
Accountability
Office

Student Questionnaire

This questionnaire should take less than five minutes. We need to know about your background, so we can see how students with different experiences do on the test. Your responses will remain completely anonymous. All the questionnaire responses will be summarized and reported at the school, board and provincial levels.

Read each question carefully. Fill in the circle that best describes your response. Make sure the circle is entirely filled in. If no answer seems exactly right, fill in the circle with the response closest to what you want to say. If you wish to change one of your answers, put an X through the incorrect answer and fill in the appropriate circle as shown below.

● Final answer   ✖ Incorrect answer

Reading

1. Indicate the types of materials you read in English outside school most weeks *(fill in all circles that apply)*.

   ○ Non-fiction books, e.g., biographies
   ○ Comics
   ○ Web sites, e-mail, chat messages
   ○ Letters
   ○ Magazines
   ○ Manuals, instructions
   ○ Newspapers
   ○ Novels, fiction, short stories
   ○ Poetry, song lyrics
   ○ Religious or spiritual writings

2. Indicate the number of hours a week you read materials written in English, not including your homework, outside school *(fill in one circle only)*.

   ○ one hour or less
   ○ more than one hour and less than three hours
   ○ more than three hours and less than five hours
   ○ five hours or more

3. Indicate what English-language materials you have at home *(fill in all circles that apply)*.

   ○ Dictionaries, encyclopedias (print or electronic)
   ○ Books
   ○ Newspapers
   ○ Magazines

Writing

4. Indicate the types of writing you do in English outside school most weeks *(fill in all circles that apply)*.

   ○ E-mail messages, chat-room conversations
   ○ Letters, journals, diaries
   ○ Notes, directions, instructions
   ○ Song lyrics, poems
   ○ Stories, fiction
   ○ Work-related writing

5. Indicate the number of hours a week you write in English, not including your homework, outside school *(fill in one circle only)*.

   ○ one hour or less
   ○ more than one hour and less than three hours
   ○ more than three hours and less than five hours
   ○ five hours or more

Home Computer Use

6. Indicate how often you use a computer at home for school work *(fill in one circle only)*.

   ○ I don't have a computer at home.
   ○ I **never or hardly ever** use the computer for school work.
   ○ I use the computer **once or twice a month** for school work.
   ○ I use the computer **once or twice a week** for school work.
   ○ I use the computer **almost every day** for school work.

Language

7. a) Is English the first language you learned at home?
      ○ Yes
      ○ No

   b) What languages do you speak at home *(fill in one circle only)*?
      ○ Only or mostly English
      ○ Another language (or languages) as often as English
      ○ Only or mostly another language (or other languages)

Thank you for completing this questionnaire. Please return it to the teacher administering the test.

Appendix II: EQUIP S.E.S Variable View

Variables Involved in the Income Index
p_be_20: proportion of households: annual income<$20,000
p_20_39: proportion of households: annual income $20,000-$39,999
p_40_59: proportion of households: annual income $40,000-59,999
p_60_79: proportion of households: annual income 60,000-79,999
p_80_99: proportion of households: annual income 80,000-$99,999
p_g_100: proportion of households: annual income $100,000 + up

Variables Involved in the Education Index
p_b_g9: proportion of pop: < G9 education
p_g9_bhs: proportion of pop: > G9 education but incomplete high school
p_hi_sc: proportion of pop: complete high school
p_ps_sec: proportion of pop: some post-secondary
p_bach: proportion of pop: at least bachelor degree

Appendix III: Exploratory Factor Analysis Outputs

(1)

7-Factor Matrix   Total Variance Explained

Compone nt	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.867	16.814	16.814	3.867	16.814	16.814	**1.939**	**8.430**	**8.430**
2	1.652	7.181	23.995	1.652	7.181	23.995	**1.876**	**8.155**	**16.585**
3	1.518	6.598	30.593	1.518	6.598	30.593	**1.726**	**7.505**	**24.089**
4	1.313	5.710	36.303	1.313	5.710	36.303	**1.585**	**6.892**	**30.981**
5	1.205	5.241	41.544	1.205	5.241	41.544	**1.543**	**6.708**	**37.689**
6	1.078	4.686	46.230	1.078	4.686	46.230	**1.533**	**6.664**	**44.352**
7	1.016	4.416	50.646	1.016	4.416	50.646	**1.447**	**6.293**	**50.646**
8	.963	4.189	54.834						
9	.910	3.956	58.790						
10	.874	3.798	62.588						
11	.848	3.686	66.274						
12	.836	3.636	69.910						
13	.809	3.518	73.429						
14	.779	3.385	76.814						
15	.759	3.300	80.114						
16	.706	3.070	83.184						
17	.672	2.923	86.107						
18	.637	2.769	88.877						
19	.600	2.608	91.484						
20	.536	2.332	93.816						
21	.518	2.251	96.067						
22	.472	2.052	98.119						
23	.433	1.881	100.000						

Extraction Method: Principal Component Analysis.

Rotated Component Matrix(a)

	Component						
	1	2	3	4	5	6	7
1.TypesOfReadingNonFiction	-.057	**.517**	-.080	.106	.260	.050	.042
2.TypesOfReadingComics	.204	.281	.073	.218	.043	-.152	.140
3.TypesOfReadingInternet	.062	.109	**.752**	.059	-.044	.022	.135
4.TypesOfReadingLetters	.152	.238	.110	.133	.015	.027	**.711**
5.TypesOfReadingMagazines	**.714**	.014	.014	.217	.005	-.017	.132
6.TypesOfReadingManuals	.055	**.653**	.219	.019	-.068	.017	.094
7.TypesOfReadingNewspapers	**.436**	.244	-.034	-.198	.109	.151	.262
8.TypesOfReadingNovels	.008	.097	.035	-.049	**.757**	.067	.113
9.TypesOfReadingPoetry	.083	.123	.131	**.759**	.077	.054	.109
10.TypesOfReadingReligious	-.028	**.511**	.006	.205	.049	-.053	.283
11.HoursReadingPerWeek	.043	.124	.066	-.019	.120	**.795**	.048
12.ReadingMaterialDictionaries	.136	.222	**.376**	.039	**.351**	-.055	-.071
13.ReadingMaterialbooks	.278	.089	.110	.050	**.551**	.021	.010
14.ReadingMaterialNewspapers	**.606**	.175	.151	-.094	.193	.073	.077
15.ReadingMaterialMagazines	**.752**	.050	.105	.247	.021	.003	-.002
16.TypesOfWritingEmail	.123	-.048	**.785**	.057	-.049	-.020	.040
17.TypesOfWritingLetters	.130	-.010	-.021	.068	.217	.113	**.762**
18.TypesOfWritingNotes	.165	**.480**	.051	-.053	.126	.137	.073
19.TypesOfWritingSongs	.136	.021	.011	**.802**	.053	.061	.054
20.TypesOfWritingStories	-.014	.138	-.146	.233	**.508**	.077	.215
21.TypesOfWritingWork	.211	**.569**	-.100	-.024	.109	.139	-.147
22.HoursWritingPerWeek	.053	.072	.021	.118	-.016	**.826**	.100
23.HomeComputerUse	-.037	-.044	**.475**	.029	.239	.296	-.143

Extraction Method: Principal Component Analysis.   Rotation Method: Varimax with Kaiser Normalization.

a   Rotation converged in 6 iterations.

(2)

**3-Factor Matrix   Total Variance Explained**

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.867	16.814	16.814	3.867	16.814	16.814	**2.655**	**11.544**	**11.544**
2	1.652	7.181	23.995	1.652	7.181	23.995	**2.534**	**11.019**	**22.563**
3	1.518	6.598	30.593	1.518	6.598	30.593	**1.847**	**8.030**	**30.593**
4	1.313	5.710	36.303						
5	1.205	5.241	41.544						
6	1.078	4.686	46.230						
7	1.016	4.416	50.646						
8	.963	4.189	54.834						
9	.910	3.956	58.790						
10	.874	3.798	62.588						
11	.848	3.686	66.274						
12	.836	3.636	69.910						
13	.809	3.518	73.429						
14	.779	3.385	76.814						
15	.759	3.300	80.114						
16	.706	3.070	83.184						
17	.672	2.923	86.107						
18	.637	2.769	88.877						
19	.600	2.608	91.484						
20	.536	2.332	93.816						
21	.518	2.251	96.067						
22	.472	2.052	98.119						
23	.433	1.881	100.000						

Extraction Method: Principal Component Analysis.

Rotated Component Matrix(a)

	Component		
	1	2	3
TypesOfReadingNonFiction	**.466**	.120	-.091
TypesOfReadingComics	.124	**.414**	.018
TypesOfReadingInternet	-.023	.203	**.683**
TypesOfReadingLetters	.335	**.466**	.016
TypesOfReadingMagazines	.073	**.609**	.106
TypesOfReadingManuals	**.341**	.213	.177
TypesOfReadingNewspapers	**.404**	.223	.058
TypesOfReadingNovels	**.516**	.027	.056
TypesOfReadingPoetry	.092	**.558**	.095
TypesOfReadingReligious	**.344**	.321	-.085
HoursReadingPerWeek	**.575**	-.219	.301
ReadingMaterialDictionaries	.211	.182	**.365**
ReadingMaterialbooks	**.353**	.228	.165
ReadingMaterialNewspapers	.294	**.355**	.263
ReadingMaterialMagazines	.063	**.610**	.221
TypesOfWritingEmail	-.164	.193	**.731**
TypesOfWritingLetters	.383	.346	-.074
TypesOfWritingNotes	**.443**	.140	.097
TypesOfWritingSongs	.029	**.562**	.005
TypesOfWritingStories	**.444**	.200	-.150
TypesOfWritingWork	**.449**	.108	-.005
HoursWritingPerWeek	**.501**	-.140	.255
HomeComputerUse	.146	-.119	**.548**

Extraction Method: Principal Component Analysis.   Rotation Method: Varimax with Kaiser Normalization.

a   Rotation converged in 7 iterations.

Scree Plot
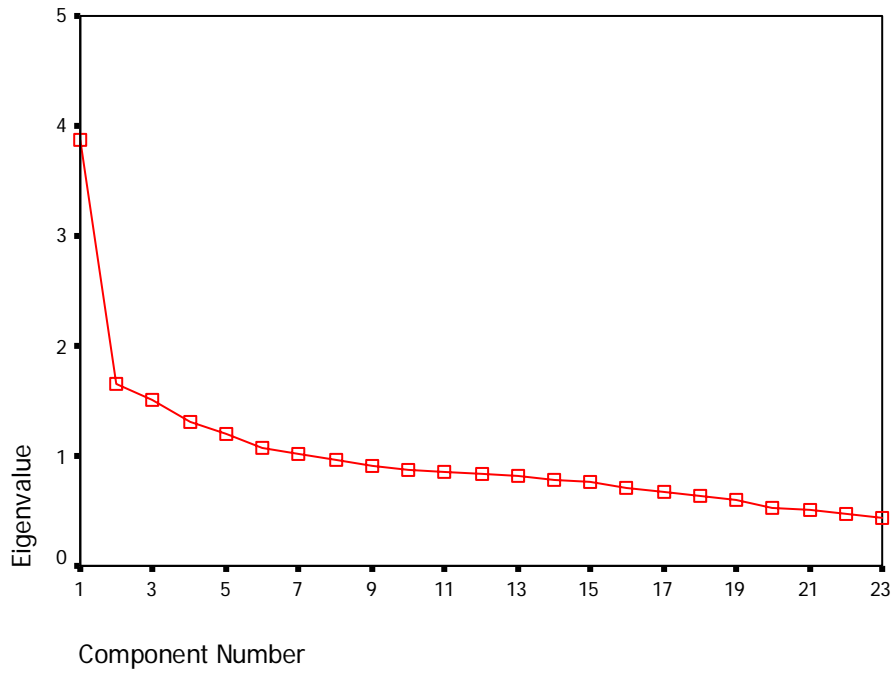
Appendix IV: Correlations among 7 Factors

	e-literacy	Literature literacy	Non-fiction literacy	Newspaper and magazine literacy	Literacy hours	Letter reading and writing	Song and poetry literacy
e-literacy	1	.073(**)	.101(**)	.157(**)	.171(**)	.074(**)	.126(**)
Literature literacy		1	.305(**)	.221(**)	.186(**)	.258(**)	.155(**)
Non-fiction literacy			1	.347(**)	.216(**)	.311(**)	.200(**)
Newspaper and magazine literacy				1	.154(**)	.335(**)	.269(**)
Literacy hours					1	.176(**)	.110(**)
Letter reading and writing						1	.269(**)
Song and poetry literacy							1

Appendix V: Analysis of Home Language Patterns

Independent Samples Test of First Language

	Levene's Test for Equality of Variances		t-test for Equality of Means		
	F	Sig.	t	df	Sig. (2-tailed)
Reading	10.841	.01	-7.575	4176	.000
Writing	14.839	.000	-7.251	4176	.000

ANOVA of Language Spoken at Home

		Sum of Squares	df	Mean Square	F	Sig.
Reading Total Score	Between Groups	70663.885	2	35331.942	32.212	.000
	Within Groups	4565092.928	4162	1096.851		
	Total	4635756.812	4164			
Writing Total Weighted Score	Between Groups	89947.495	2	44973.748	30.628	.000
	Within Groups	6111338.195	4162	1468.366		
	Total	6201285.690	4164			

* The mean difference is significant at the .01 level.

Post hoc tests: Multiple Comparisons of Language Spoken at Home

Dependent Variable	(I) Languages At Home	(J) Languages At Home	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Reading Total Score	only or mostly English	another language (or languages) as often as English	-1.45	2.274	1.000	-6.90	3.99
		only or mostly another language (or other languages)	-9.61(*)	2.202	.000	-14.88	-4.33
	another language (or languages) as often as English	only or mostly English	1.45	2.274	1.000	-3.99	6.90
		only or mostly another language (or other languages)	-8.15(*)	1.098	.000	-10.78	-5.52
	only or mostly another language (or other languages)	only or mostly English	9.61(*)	2.202	.000	4.33	14.88
		another language (or languages) as often as English	8.15(*)	1.098	.000	5.52	10.78
Writing Total Weighted Score	only or mostly English	another language (or languages) as often as English	-6.64(*)	2.631	.035	-12.94	-.34
		only or mostly another language (or other languages)	-14.62(*)	2.547	.000	-20.72	-8.52
	another language (or languages) as often as English	only or mostly English	6.64(*)	2.631	.035	.34	12.94
		only or mostly another language (or other languages)	-7.98(*)	1.271	.000	-11.02	-4.94
	only or mostly another language (or other languages)	only or mostly English	14.62(*)	2.547	.000	8.52	20.72
		another language (or languages) as often as English	7.98(*)	1.271	.000	4.94	11.02

Appendix VI: Ethical Approval for the Study

Queen's
UNIVERSITY

OFFICE OF RESEARCH SERVICES

December 9, 2004

Fleming Hall, Jemmett Wing
Queen's University
Kingston, Ontario, Canada K7L 3N6
Tel 613 533-6081
Fax 613 533-6806
ors@post.queensu.ca
www.queensu.ca/vpr/

Ying Zheng
Graduate Student
Faculty of Education
Queen's University

GREB Ref #   GEDUC-184-04
Title: "Exploring Factors Associated with ESL/ELD Students' Performance in the Ontario Secondary School Literacy Test"

Dear Ying Zheng:

The General Research Ethics Board (GREB) has given expedited approval to your proposal entitled **"Exploring Factors Associated with ESL/ELD Students' Performance in the Ontario Secondary School Literacy Test"**. In accordance with the Tri-Council Guidelines (article D.1.6) and Senate Terms of Reference (article G), your project has been approved for one year. At the end of each year, GREB will ask if your project has been completed and if not, what changes have occurred or will occur in the next year.

You are reminded of your obligation to advise the GREB, with a copy to the E-REB, of any adverse event(s) that occur during this approval period (form is attached). An adverse event includes, but is not limited to, a complaint, a change or unexpected event that alters the level of risk for the researcher or participants or situation that requires a substantial change in approach to a participant(s). You are also advised that any adverse events must be reported to the GREB within 48 hours.

You are also reminded that all changes that might affect human participants must be approved by the GREB. Examples of required approvals are: changes in study procedures or implementations of new aspects into the study procedures that affect human subjects. These changes must be sent to Linda Frid at the Office of Research Services or fridl@post.queensu.ca prior to implementation. Ms. Frid will seek the approval of the GREB reviewer(s) who originally assessed your application.

On behalf of the General Research Ethics Board, I wish you continued success in your research.

Yours sincerely,

Lee Fabrigar, PhD
Associate Professor and
Member, General Research Ethics Board

LF/lf
c.c.:   Ruth Rees, Chair of E-REB
         Liying Cheng, Faculty Supervisor
         Heather Cross

Attach: Adverse Events Form

think Research
think Queen's

PREPARING LEADERS AND CITIZENS FOR A GLOBAL SOCIETY

Appendix VII: Sample OSSLT Booklet

Get Rid of that T-shirt!

A recent newspaper article pointed out that Canadians purchased 73.7 million T-shirts last year. The article went on to say that the average North American owns 25 of them. The T-shirt was praised as the favourite garment of the twentieth century, worn by men and women, young and old, rich and poor. As we begin a new century, I suggest we leave the old T-shirt behind.

The first wearers of an "undershirt" or a "work shirt" in public were making a rebellious statement, but it quickly became the accepted style. Eventually, we all began to wear underwear anywhere and everywhere.

In the 60s, hippies tie-dyed their T-shirts. In the 70s, punk rockers shredded, safety-pinned and spray-painted them. In the 80s, T-shirts became great democratic portable billboards — each shirt an editorial column or personal ad telling others about the places the wearer has been, or the products, bands and politics the wearer supports or abhors.

The most recent trend seems to be toward slogans or messages that are increasingly meaningless. The best known examples are expensive T-shirts sporting only the name of the manufacturer. It seems strange that people are now expressing themselves by broadcasting their support of a shirt manufacturer. I can't think of anything less individualistic or less attractive to wear in public.

The T-shirt is basically a formless, ugly garment. What should happen to the 25 T-shirts each of us is supposed to have? I suggest that we use them as rags for washing our 1.7 cars.

Reading 5

multiple choice (Circle the letter next to the best or most correct answer for each question.)

1. In this selection, the T-shirt is compared to
   - A   a slogan.
   - B   a product.
   - C   a garment.
   - D   a billboard.

2. Which of the following is the best way to describe the purpose of this selection?
   - F   to state an opinion
   - G   to describe a product
   - H   to present information
   - J   to provide instructions

written answers

3. What is the meaning of the phrase "broadcasting their support of a shirt manufacturer" as used in paragraph 4?

   _____

   _____

4. Explain the purpose of the question in paragraph 5.

   _____

5. Do you think T-shirts will continue to be popular? Use one piece of information from this selection to support your answer.

   _____

Writing a Summary

Get Rid of that T-shirt!

Rough Notes

A recent newspaper article pointed out that Canadians purchased 73.7 million T-shirts last year. The article went on to say that the average North American owns 25 of them. The T-shirt was praised as the favourite garment of the twentieth century, worn by men and women, young and old, rich and poor. As we begin a new century, I suggest we leave the old T-shirt behind.

The first wearers of an "undershirt" or a "work shirt" in public were making a rebellious statement, but it quickly became the accepted style. Eventually, we all began to wear underwear anywhere and everywhere.

In the 60s, hippies tie-dyed their T-shirts. In the 70s, punk rockers shredded, safety-pinned and spray-painted them. In the 80s, T-shirts became great democratic portable billboards — each shirt an editorial column or personal ad telling others about the places the wearer has been, or the products, bands and politics the wearer supports or abhors.

The most recent trend seems to be toward slogans or messages that are increasingly meaningless. The best known examples are expensive T-shirts sporting only the name of the manufacturer. It seems strange that people are now expressing themselves by broadcasting their support of a shirt manufacturer. I can't think of anything less individualistic or less attractive to wear in public.

The T-shirt is basically a formless, ugly garment. What should happen to the 25 T-shirts each of us is supposed to have? I suggest that we use them as rags for washing our 1.7 cars.