



Project  
**MUSE**<sup>®</sup>  
*Scholarly journals online*

# Student Accounts of the Ontario Secondary School Literacy Test: A Case for Validation

---

Liyong Cheng  
Janna Fox  
Ying Zheng

**Abstract:** The Ontario Secondary School Literacy Test (OSSLT) is a cross-curricular literacy test issued to all secondary school students in the province of Ontario. The test consists of a reading and a writing component, both of which must be successfully completed for secondary school graduation in Ontario. This study elicited 16 first language and second language student accounts of their OSSLT test-taking processes immediately after the March 2006 test administration. The analysis of these students' accounts provided valuable information about the validity of the inferences drawn from the Ontario Secondary School Literacy Test. These accounts suggest the complexity of the processes that the students engaged as they attempted to demonstrate their reading and writing skills on the test. The study has implications for test developers and test users regarding the interpretation of student test performance on the Ontario Secondary School Literacy Test.

**Keywords:** validation; student accounts; large-scale testing; reading and writing; second language students

**Résumé :** Le *Test provincial de compétences linguistiques (TPCL)* est un test de littératie transdisciplinaire soumis à tous les élèves des écoles secondaires de l'Ontario. Le TPCL comporte une épreuve de lecture et une épreuve d'écriture qui doivent être réussies pour obtenir le diplôme d'étude secondaire de l'Ontario. L'étude a permis de recueillir des comptes rendus auprès de 16 élèves de langue maternelle (L1) et de langue seconde (L2) sur les processus utilisés pendant le test, immédiatement après la passation de mars 2006. L'analyse de ces comptes rendus a fourni une information utile quant à la validité des inférences à partir du TPCL. Les comptes rendus traduisent la complexité des processus mis en œuvre alors que les élèves tentent de démontrer leurs habiletés en lecture et en écriture à travers le test. L'étude a des implications pour ceux qui développent le test et les utilisateurs du test en ce qui a trait à l'interprétation de la performance des élèves au TPCL.

**Mots clés :** validation; verbalisations des élèves; testing à grande échelle; lecture et écriture; élèves en langue seconde

## Introduction

English as a second language (ESL) and English literacy development (ELD)<sup>1</sup> students are estimated to form 20 to 50% of the general student population in urban K–12 systems across Canada (Roessingh, 1999). Ontario secondary schools have seen the greatest increase of ESL/ELD students over the past 20 years. The Ontario curriculum defines ESL students as students who are learning English, can read and write in their first language (L1), and who mostly have had continued schooling before arriving in Canada. ELD students, on the other hand, are those who do not read and/or write in their first language (L1), i.e., they are learning to read and write in English, in their second/third language, and they may have missed years of schooling (Ministry of Education and Training, 1999).

At the same time, large-scale achievement testing has increasingly been used to measure and ensure student competency or provide system accountability (Firestone, Mayrowetz, & Fairman, 1998; Ryan, 2002). As evidenced by ESL/ELD students' low success and high deferral rates on the Ontario Secondary School Literacy Test (OSSLT) (Education Quality and Accountability Office [EQAO], 2002, 2003, 2004), many of these students are struggling as they attempt to obtain fundamental English literacy skills in their schooling and pass the OSSLT. For example, in October 2003, only 42% of the ESL/ELD students passed the whole test as compared to an overall pass rate of 77%. About 45% of the ESL/ELD students passed reading only, and about 69% of the students passed writing only (as compared to the overall rates of 82% and 88% respectively for all students who wrote the test). In terms of deferral rates, just over half of the ESL/ELD students (54%) participated in the test as compared to an overall participation rate of 91% (EQAO, 2003).

The confluence of increased numbers of ESL/ELD students in Ontario and this expanding testing framework has created a new and largely unanticipated educational problem – alarmingly high failure rates for these students (Watt & Roessingh, 2001). In addition, large-scale tests such as the OSSLT are constructed for first language English students. Research suggests, however, that they may have lower reliability and validity for second language students and should be interpreted differently (Abedi, Leon, & Mirocha, 2003). Clearly, tests like the OSSLT add extra challenges to the academic success of these L2 students. There is an urgent need to investigate what the test means to L2 students and whether the test has different meanings for different groups of test-takers. Therefore, we designed a

three-phase study to investigate the impact of the OSSLT on L2 students in comparison with their L1 counterparts. The first phase of the study investigated the impact of the test formats, test constructs (see Cheng, Klinger, & Zheng, 2007), and home language, computer use, and after-school literacy activities on these students (Klinger, Cheng, & Zheng, under review). The second phase, which is reported in the present article, aimed to understand the differences in L1 and L2 students' accounts of the test. This information may inform the validity of criteria and procedures in the OSSLT and help us understand why these L2 students fail at a much higher rate than non-ESL/ELD students. In phase two, we also examined how literacy was defined in the test in relation to how students used reading and writing for academic purposes in schools (see Fox & Cheng, 2007). A large-scale survey was conducted in phase three. Questionnaires, developed on the basis of findings from phases one and two, were administered in the fall of 2006 to Grade 10 (pre-test) and Grade 11 (post-test) students across secondary schools in Ontario to explore differences in test performance and perceptions of the test across L1 and L2 groups.

Previous research has also suggested the reading and writing performance of L2 students could be influenced by their personal characteristics. For instance, students who are literate in their L1 generally make faster progress in L2 acquisition (Bernhardt & Kamil, 1995; O'Malley & Chamot, 1990) and do better on tests (Cummins, 1982, 1983) than those without L1 literacy, i.e., ELD students. Length of stay in Canada in relation to the age when students are immersed in English (Collier, 1987, 1988; Cummins, 1981; DiCerbo, 2000; Roessingh & Kover, 2003) and reading/writing at home (Collier, 1992; Krashen, 1987, 2007) have also been shown to have a significant impact on academic learning and success. In addition, research has indicated that cultural capital, including prior knowledge and language skills, acts as a filter in disadvantaging L2 students who do not possess English language proficiency (Garcia & Gopal, 2003). Thus, a closer look at the experiences with the OSSLT reading and writing performance of the ESL/ELD group could identify subgroups that are most at risk.

### **Test validation and student accounts**

Achieving validity in tests like the OSSLT is conceptually challenging and difficult to establish (Cumming, 1996, Solano-Flores & Trumbull, 2003). Psychometric and mentalistic approaches are complementary means of accruing evidence of test validation. Statistical aspects of

tests, following psychometric approaches to validity evidence, are the focus of the vast majority of validation studies. However, the psychometric approach cannot access certain properties or features of test performance. For example, what test designers think they are testing may not be the same as what test-takers think the test is measuring. Correct answers to a question can be arrived at through a correct hypothesis, an incorrect one (Nevo, 1989), or wild guessing. An understanding of these kinds of processes can be accessed through the mentalistic approach, offering direct evidence of test-takers' reasoning with regard to test items. Judgments can be made based on verbal reports (student accounts) of whether or not the test answers are made by good or bad reasoning (Norris, 1991).

Protocol analysis employs test-takers' verbal reports to access covert mental processing (Ericsson & Simon, 1987). Verbal report data have been used to fulfil several assessment functions (Banerjee & Luoma, 1997), which include but are not limited to exploring the processes and strategies test-takers use to complete a task, to refining the test instrument, or to refining feedback procedures. Kunnan (1998) reviewed language assessment studies using Messick's validity framework. Most studies using protocol analysis to investigate test validity fall into the two categories of 'test-taking processes' and 'test-taker strategies.' Such investigations are intertwined with other aspects of test validation. For example, some studies have focused more on students' processing of a certain type of test (Crain-Thoreson, Lippman, & McClendon-Magnuson, 1997; Ericsson & Simon, 1987; Farr, Pritchard, & Smitten, 1990; Sasaki, 2000; Sepassi, 2003; Storey, 1997), and others focused more on strategies involved in solving language problems (Cohen, 1997; Garner, 1988; Kuusela & Paul, 2000; Wijgh, 1996).

There are two broad categories of these kinds of studies – studies that focus on particular skills (e.g., writing, reading comprehension), and studies that explore cultural differences in problem-solving (Duran, 1989), the effects of cultural schemata (Parry, 1987; Pritchard, 1990), acculturational factors (Ortiz, 2005), cultural validity (Solano-Flores & Nelson-Barber, 2001), background knowledge (Clapham, 1998; Ginther & Stevens, 1998; Jennings, Fox, Graves, & Shohamy, 1999), and 'situatedness' (Atkinson, 2005, pp. 50–51).

Writing and composition research has a long-standing tradition of using verbal protocol analysis that can be traced back to cognitive approaches to writing and a focus on writing processes (see, for example, Ericsson & Simon, 1984; Hayes & Flower, 1980; Scardamalia & Bereiter, 1987). In fact, many of these early writing

researchers were cognitive psychologists. The use of verbal reports as a data elicitation technique has occasioned vociferous debate within the field of writing and composition studies over the years (see, for example, Bracewell, 1994; Cooper & Holzman, 1983; Smagorinsky, 1991). There are many critics of verbal reports. Some argue that writers are unaware of what they do as writers, may not have the ability to describe what they are doing because of the complexity of the behaviour (Cavanaugh & Perlmutter, 1982), or may fail to mention in their verbal reports what is key to understanding their writing processes. Others argue that there is no way of ensuring that writers will report what they actually did and thus the genuineness of their reports is always an issue (Nisbett & Wilson, 1977). Further, writers may report what they think researchers would like to hear. Although verbal report data continues to be viewed by most writing researchers as a useful research tool (Cumming, 1990; Manchon, Murphy, & Roca de Larios, 2005; Smagorinsky, 1991), it is important to interpret these results with caution.

Studies that have examined reading comprehension using protocol analysis are also rich in the literature (e.g., Cohen, 1984, 2007; Crain-Thoreson et al., 1997; Nevo, 1989; Phakiti, 2003; Sasaki, 2000; Storey, 1997; Wijgh, 1996). Crain-Thoreson et al. addressed methodological questions about the use of think-aloud report procedures and the roles of prior knowledge and strategy use in reading comprehension. They concluded that verbal reports can provide a useful and valid window on reading comprehension processes, at least with skilled adult readers, but there were still possibilities that the think-aloud process could facilitate or disrupt comprehension, depending on the subjects' proficiency levels or the difficulty levels of the reading passages. Cohen's (1984) study investigated 'the fit between test constructors' presumption about what is being tested and the actual processes that the test taker goes through' (p. 70). He commented, 'Due to the flaws in the test or due to certain test-taking strategies, students may not be displaying a representative performance of their language competence' (p. 71). These results imply that test constructors should make sure that the instructions for a test reflect the most efficient approach to taking that particular test.

Nevo (1989) investigated the strategies activated while processing a test in the first language and in the target language. The study revealed that various strategies are employed by respondents in taking a multiple-choice reading comprehension test in L1 and L2. The verbal reports from the subjects also revealed that items with high linguistic demands and low degrees of explicitness led to processing difficulty.

Storey (1997) pointed out that to validate a test, two fundamental aspects need to be examined. One is the test's representativeness of the selected domain – construct representativeness as in Messick (1989). The other is the process that is engaged as test-takers write the test – a crucial and often problematic issue in validation. Storey argued that if an item generated processes that coincided with those that the theory allowed (i.e., based on an identified theoretical model), the construct validity of that item/test could be claimed. If, on the other hand, construct irrelevance was identified, a lack of construct validity was indicated. Wijgh (1996) confirmed that there were some discrepancies in the reading behaviour/strategies demonstrated in the testing situation and those hypothesized by the reading skill constructs. For example, students typically used the same basic strategies and did not orient themselves to the various difficulty levels. Only students with higher proficiency levels managed to use extra-linguistic information to compensate for their weakness in understanding (e.g., vocabulary).

In sum, in validating reading tests, protocol analysis via participants' verbal reports has provided a useful vehicle for researchers to explore students' accounts of how they proceed and what leads them to get what they get, and has thus opened a window for the comparison of students' accounts with test constructors' expectations of the intended or idealized processes in tests (e.g., Storey, 1997; Wijgh, 1996). In addition, information on students' utilization of test-taking strategies can also be evaluated apart from the more traditional means of using strategy questionnaires to collect this sort of information (e.g., Cohen, 2007; Nevo, 1989).

As mentioned above, another important branch of studies using protocol analysis involves the investigation of cultural differences in problem solving, the effects of cultural schemata, acculturational factors, cultural validity, situatedness, and background knowledge. Overall, these studies indicate that factors related to language development and acculturation are integrated with students' achievement in school, and they play an important role in explaining why certain linguistic and cultural groups of students perform differently on tests. Duran (1989) pointed out that differences in cultural background could influence the ways in which individuals perceive problem-solving situations. Linguistic minority students might be adversely affected on their test performance as a result of differences in cultural background as well as in their problem-solving styles. It is necessary to systematically study the test content and formats of items to investigate how test-takers' performances on test contents and items interact with their



language and cultural backgrounds. As Ortner (1984, cited in Atkinson, 2005, p. 51) noted, 'Studying culture *from the actor's point of view* . . . does not imply that we must get *into people's heads*. What it means, very simply, is that culture is a product of acting social beings trying to make sense of the work in which they find themselves, and if we are to make sense of a culture, we must situate ourselves in the position from which it was constructed.' Verbal protocol analysis allows us to access how test-takers *make sense of a test* in order that we may better understand if a test is eliciting the behaviours and performances we intend to measure.

Pritchard (1990) utilized think-aloud protocols to examine the strategies that readers use while reading culturally familiar and unfamiliar materials. Students recalled significantly more idea units when they were reading culturally familiar texts than when reading unfamiliar passages. Students reported using more strategies when reading culturally unfamiliar texts because of their lack of sufficient background knowledge. Similarly, Sasaki (2000) investigated how schemata activated by culturally familiar words/texts might have influenced students' cloze test-taking processes and found those who read culturally familiar cloze texts demonstrated better understanding of the text and performed better.

The above research literature demonstrates the usefulness of approaching test validation from the perspective of how test-takers from diverse linguistic or cultural backgrounds account for their own performances on a test. Of particular interest to the present study are the accounts of test-taking processes engaged in test performance by L1 and L2 test-takers from different cultural and linguistic as well as schooling backgrounds. Do they solve the problems posed by the OSSLT in similar ways? Do they respond to the test tasks and items in the same ways?

## Methods

The purpose of the current study was to investigate the test-taking processes of L1 and L2 students on the OSSLT. We explored the relationship between these culturally and linguistically diverse test-takers' accounts of test performance and validity criteria by eliciting and analyzing retrospective verbal reports. Retrospective verbal reports are occasional and immediate retrospective accounts (Selinger & Shohamy, 1989). These differ from concurrent/introspective verbal reports, which aim to get at mental processing in that students talk about what they do while they are doing it.



Both methods collect 'data from test-takers and/or examiners in which they talk about their thought processes while they take a test or assess a test performance' (Banerjee, 2004, p. 77), and both could serve the purpose of our study (see also Gosden, 1996). We chose retrospective verbal reports as the method of the data collection for reasons of ethicality and practicality. Our data were collected immediately after the test and/or within a day after the test.

### *Participants*

Participants were 16 students from two schools of which nine were L2 (ESL/ELD) students and seven were L1 (non-ESL/ELD) students. They were invited via their school and they agreed to participate in the study with their parents' consent. Both schools (we named them School A and School B) have high ESL populations. According to the March 2006 OSSLT report, 52% of School A's and 30% of School B's school population used their first language at home rather than English (see <http://www.eqao.com>). The nine L2 students were all born outside Canada and had been in Canada from three months to about two years at the time of the test in March 2006. They spoke at least one language other than English. Among the L2 group, four did not pass the March 2006 OSSLT. All four students had been in Canada for only about six months by the time they took the test (all were first-time eligible test-takers, i.e., they took the test for the first time). On the other hand, five out of the seven L1 group were born in Canada and spoke English as a first language; two were born outside of Canada but had been in Canada for the past five years. The choice to take or defer the test rested with the students and the recommendation of their ESL teachers. Sometimes students were recommended to take the test even though they felt they might not pass the test, in order to gain familiarity with it (personal communication with an ESL teacher at School A, February 2006).

### *The test*

The OSSLT is a provincial test that assesses students' reading and writing skills that have been acquired by Grade 10. It is based on the reading and writing skills defined in the Ontario Curriculum across all subject areas up to the end of Grade 9. The Education Quality and Accountability Office (EQAO) developed and administers the OSSLT as a quality assurance measure to provide evidence that all Ontario secondary school graduates have met a common,

minimum standard for literacy.<sup>2</sup> Since its introduction as a high school graduation requirement in 2002, there have been five administrations of the test (including the March 2006 test administration, which was the focus of the present study). Students are eligible to write the test in Grade 10, although they may defer writing the test if they wish. They may rewrite the test if they fail on their first attempt, and if unsuccessful may take a literacy course, which if passed will also satisfy the graduation requirement.

The test is divided into separate reading and writing tasks, with three hours allotted for the completion of the test. L2 students are allowed additional time as an accommodation and may take as much time as they need to complete the test. The test is administered in two 75-minute sessions using two booklets. Each booklet consists of:

- reading selections with multiple-choice and open-response questions,
- multiple-choice writing-related questions,
- short writing prompts, and
- a long writing prompt.

The reading test has multiple-choice and open-response questions that focus on the reading skills 'required in school and daily life' (Education Quality and Accountability Office, 2006, p. 3). The reading passages on the test include an information paragraph, a news report, dialogue, real-life narrative, and graphic texts (e.g., a schedule). They vary in length from a single paragraph to two pages. The writing test has short and long writing tasks, and multiple-choice questions that focus on three writing skills 'required in school and daily life': developing a main idea with sufficient supporting details, organizing information and ideas in a coherent manner, and using conventions (spelling, grammar, punctuation) in a manner that does not distract from clear communication. There are long writing tasks that include a news report (one page) and a series of paragraphs expressing an opinion on a given topic (two pages); short writing tasks (six lines each); and multiple-choice questions on organization of ideas, sentences and paragraph structure, and language conventions.

#### *Data collection and analysis*

We used focus groups immediately after the March 2006 OSSLT administration as the means to collect students' retrospective verbal reports. The rationale for focus groups was to enable understanding

from a group point of view of students' accounts of what the OSSLT measured. Focus groups are fundamentally a way of listening to participants and learning from them as a group (Krueger & Casey, 2000) as they share a common experience. In this study, their common experience was the OSSLT they had just completed. The focus group method had advantages over individual interviews for this study because it created an opportunity for participants to interact with small groups of their peers and to discuss and share their thoughts, feelings, attitudes, and ideas about their experiences with the OSSLT.

Participants took part in the focus groups immediately after the March OSSLT administration, either on March 29 (the same day as the test), or one day after the test on March 30, 2006. They participated in two separate focus groups – one on the reading and one on the writing parts of the test. Each focus group was conducted with two to three students and moderated by one researcher. L1 and L2 students were grouped separately. In total, four researchers were involved with the focus groups.

Focus groups began with the researchers explaining the purpose of the study and eliciting general information on the participants' first language backgrounds, their general responses to the OSSLT, and their overall test-taking experience (see Appendix 1 for the focus group questions). Then questions were asked about their reactions to the reading and writing tasks (i.e., which kind of tasks were more challenging for them). Attention was then directed to the two narrative reading tasks and the news report writing task, based on our results from the first phase of the study on the same test (i.e., the 2002 and 2003 OSSLT administrations; see Cheng et al., 2007). In that study, findings based on discriminant analyses of students' test performance on each facet of their reading constructs/formats and writing tasks indicated that the narrative reading and the news report tasks most distinctively separated the L2 (ESL/ELD) and L1 (non-ESL/ELD) groups in their test performance. Students' accounts of their test-taking process, reported below, allowed for cross-referencing (triangulation) of these findings regarding the impact of this test on L2 students (see also Fox & Cheng, 2007).

We were able to use the same test items/tasks that students wrote on the 'live' test for our focus groups. Each of the focus groups (reading or writing) took about 20 minutes to complete. Altogether, 21 focus groups<sup>3</sup> were conducted, of which 12 were with L2 students and 9 were with L1 students. All focus groups were audio-recorded and transcribed for analysis, using a grounded theory approach

(Strauss & Corbin, 1990) in which the sorting, coding, and comparison of data arising from the focus groups of L1 and L2 test-takers resulted in the identification of key characteristics of their test-taking processes.<sup>4</sup> Subsequently, these categories were reduced to those representing key or subordinate differences as we re-examined and discussed our coded transcripts. The categories of key differences in L1 and L2 test-taker responses to the OSSLT are discussed below.

## Results and discussion

Five key categories were derived from the data as the main findings of our study. They are general test (task) difficulty, test format, test-taking strategies, test-taking processes, and test topics.

### *General test (task) difficulty*

Before specific questions were asked about how students processed each test task, we asked how they felt about the test in general. This gave us a sense of how the L1 and L2 groups perceived and approached the test.

L1 students found that the test tasks were 'simple and basic concepts'; 'pretty easy, I didn't have any problems with it' (L1 Student J, School A).<sup>5</sup> '[I]t could have at least challenged us a bit more with . . . more of a time constraint' (L1 S-N, S\_B). One of the L1 students commented, 'I liked it. Writing has never really been a problem for me. It wasn't really hard' (L1 S-N, S\_B).

However, the test provided a rather different scenario for L2 students. A majority of the L2 students mentioned the test was 'hard' and 'difficult':

So, so difficult, and not easy. (L2 S-M, S\_A)

Difficult, I don't know, I don't know, I can't remember all of the questions. (L2 S-N, S\_A)

Hard. (L2 S-Sh, S\_B)

Quite difficult, vocabulary is difficult. (L2 S-E, S\_A)

Regarding the kind of difficulties they experienced, some commented on the inferential reading questions and writing conventions. Many L2 students mentioned vocabulary as one of the key difficulties in taking the test. For example, many of the L2 students did not know the meaning of a certain word (i.e., invention) in the six-line writing task

prompt (see Appendix 1: Writing, Section G) so they had a hard time completing the task: 'Yeah. That invention. That word. I don't know what the meaning. So I don't know how to write about it' (L2 S-M, S\_A). Another L2 student, who remarked that she found vocabulary on the test difficult, also reported that it prevented her from completing the six-line writing task. She also didn't know the word invention, so she wrote about international films, which had nothing to do with the topic. Another L2 student identified the same issue: 'Some vocabulary I cannot understand. And so... I cannot make the correct answer' (L2 S-E, S\_A).

The above responses provide a general idea of how the L1 and L2 groups responded to the overall difficulty level of the test and their general experiences of the test. The responses also suggested that for many L2 students this test was a test of language proficiency/vocabulary, as a single unknown word stopped them from carrying out the writing task. This result corroborated the concern that students' vocabulary limitations can sometimes turn writing tests into tests of vocabulary (Wu, 1998), and it provided further evidence that vocabulary was a key indicator of language proficiency.

However, the difficulty of a task was not related to the amount of time allowed to do the test. As one of the test accommodations of the OSSLT, L2 students were given double the normal testing time. Many L2 students remarked that they had ample time to do the test. For example, when asked about time, one L2 student responded, 'I have enough time. Like I think it's too much' (L2 S-M, S\_A). In fact, none of L2 students used up all of the two and a half hours they were given to do the test in School A. They told us that they did not know how to do some of the tasks, but the extra time did not seem to help them. As for L1 students, many of them felt they had too much time to play with: 'Way too much time. Too much time for the test, 30 minutes left' (L1 S-N, S\_B); 'good amount of time' (L1 S-St, S\_B); 'finished over 30 minutes earlier' (L1 S-Sa, S\_B). Some of these L1 students felt that they only needed 65 to 70% of the time given for each booklet.

In addition, we found both L1 and L2 students used the word 'tired' to describe their feeling after taking the test. However, the word 'tired' was used by L1 and L2 to mean two different things. The L1 students were 'tired' because the tasks were 'tedious' and they had too much time (L1 S-St, S\_B). L2 students felt physically tired because there were too many tasks to complete. 'I have already written for hours, and then my brain is like too... too... how to say? ... I can't concentrate.' (L2 S-E, S\_A).

*Test format*

There were two narrative reading texts in the OSSLT's March 2006 administration. One was a narrative reading passage on a train crash, and the other was a dialogue narrative reading about a character called Nino and his conversation with a computer customer-service agent. When we asked specific questions about which of the narrative readings they liked or they found easy or difficult, we received different answers. However, the difference was not necessarily between L1 and L2 students. For example, there were L1 and L2 students who agreed that the dialogue narrative reading about Nino was more complex and difficult, but there were also L1 and L2 students who found the same task easy.

One L1 student said,

I thought that the Nino reading was more complex because they are sort of asking you the questions about Nino's personality . . . In the Nino one, there was multiple-choice and then you had to answer open-ended questions . . . In one of the open-ended questions, it says what kind of person is Nino. You kinds have to analyze like go back and reread what Nino says and kind of develop his personality just by reading something, which is much more difficult than just taking information and then rewriting it. (L1 S-L, S\_A)

One L2 student also commented, 'The Nino passage is more difficult, as I can't understand what they are talking about' (L2 S-K, S\_A). On the other hand, both L1 and L2 students said, 'The Nino passage is easier with straightforward answers' (L1 S-St, S\_B; L2 S-N, S\_B).

Other students had opinions different from those of the above students. For example, one L2 student commented, 'The train one is difficult. I was reading but not concentrating. I was just really bored' (L2 S-L, S\_B). Another L2 student responded similarly: 'The topic is boring but . . . one thing more is that we had to read between the lines' (L2 S-V, S\_B). Or, as another student put it, 'The train crash reading is more difficult because of a lot vocabulary. The Nino one is easier because maybe I am interested . . . I know what they are saying. They are saying about the computer' (L2 S-E, S\_A).

An L1 student also commented, 'The train crash passage is harder, as it is not a dialogue, and it has inference questions' (L1 S-H, S\_A). Although these two students agreed with each other on the difficulty of the passage, they offered different reasons. This may indicate that the difficulty of the passages could reside with the individual

students, irrespective of their English language proficiency (in terms of L1 or L2), as long as they understood the reading.

In the end, there were L1 students who commented that they found 'both readings very easy to understand and easy to answer' (L1 S-N, S\_B). And there were L2 students who commented, 'Both passages are difficult' (L2 S-S, S\_A).

For the writing tasks, L2 students continued to mention that the news report was difficult for them, which we confirmed through analysis of both L1 and L2 students' test scores (see Cheng et al., 2007). For example,

I mean the news report was harder because you have to create something, invent something out of nowhere, just from the picture. But for the opinion piece, you can just... you had your points so you had something to work with. This one [news report] you only have a picture and you have to make up something that would be realistic. (L2 S-L, S\_B)

We did not find any L1 students who mentioned the difficulty of news report writing (see also Fox & Cheng, 2007).<sup>6</sup> They, however, liked different types of writing tasks. For example, one student said, 'I liked the news report more. It left more space for creativity' (L1 S-N, S\_B). Another said, 'I really liked multiple-choice questions because they're quick and easy' (L1 S-Sa, S\_B). Other L1 students reported,

For me, I'd say the opinion piece is more demanding than the news report, ... the essay about the junk food. Yeah, I thought that was harder because for the news report I wrote it in really funny kind of a way. (L1 S-H, S\_A)

I prefer writing essays to multiple-choice writing questions because you can get into more detail with essays. You can... because it takes more... you have to actually think about yourself. With MC questions, basically, you just chose what is on the paper in the first place. (L1 S-J, S\_A)

Those [six-line writing tasks such as 'invention' and 'change'] were nice because you didn't feel like you had to write a whole lot. You just had to summarise a couple paragraphs that we had read... You didn't have to write as much as the essays. (L1 S-L, S\_A)

These different preferences for particular writing tasks echoed the varying responses we received about the reading tasks. That different students liked different types of tasks supports the selection of a range of tasks in this test.



*Test-taking strategies*

Both L1 and L2 students discussed test-taking strategies. L1 students seemed to rely on ‘more reasoning...based on text’ (L1 S-N & S-Sa, S\_B) or on processes of elimination: ‘do elimination, but do it based on information from text’ (L1 S-L, S\_A). For example, one L1 student explained his ‘reasoning’ strategy in responding to an item that was testing inference: ‘Is there evidence in the paragraph? It says, “Finally, the headlights appeared around the bend in the track” – so I was glad to see it. So it is not like another accident happening, the rescue crews came to save them. Good reasoning’ (L1 S-H, S\_A).

Another L1 student identified the same phrase in the text, in responding to the item: ‘It was kind of elimination and reading from the text, because they said “Finally headlights appeared around the bend.” So it is a little bit of reading and little bit of elimination’ (L1 S-L, S\_A)

In contrast, L2 students also used elimination, but typically focused on the multiple-choice options offered in the item rather than the source text: ‘First, I read all the information. A is not enough, B too, and then...C is too. Only D has enough information’ (L2 S-E, S\_B). This student chose the wrong answer using an elimination approach. In many cases, however, L2 students said they engaged in ‘blind’ guessing: ‘I don’t know why A and I don’t know how. I just guessed’ (L2 S-M, S\_A), or, ‘G and H and J – the word is – I couldn’t understand...I guessed H’ (L2 S-S, S\_A).

They sometimes chose an answer if they saw the exact words or phrases in both the question and the text: ‘Just...because the other answers I have a little bit vocabulary I don’t know and then...C I know and then maybe the sulphuric acid made me...to think that explode. That’s why I chose it’ (L2 S-E, S\_A).

Another distinct strategy that L1 students used was to rely on tacit understandings of words, sentences, and test choices. They repeatedly said they made a choice because it sounded right: ‘C makes the most sense. It sounds right. It is concise. It’s clear and concise and it combines the three sentences – three ideas – three sentences the best way’ (L1 S-N, S\_B). Or, ‘Like we can just read them and then just know them. We don’t have to know the grammar because we spoke a lot of it...I just read it. I think that’s what we do’ (L1 S-H, S\_A). Other examples of this tacit understanding include:

I checked each one I inserted. I just write (speak) that way. Like on top I crossed out ‘often has been called’ and which had ‘often is called’ because

they don't flow well and they don't fit. But I had actually said it 'Long Point, often is called, the Graveyard of Lake Erie' and ... but I kind of see that that's not right. (L1 S-J, S\_A)

You see the sentence structure. Then you can say it ... you'll learn ... realize what sounds wrong and what sounds right. And that's, you know, by reading a lot, by simply having conversations every day ... Listening to other people talk and learning what sounds better in speech. (L1 S-N, S\_B)

These comments are evidence that L1 students have a better sense of the language through their life experiences of it. They have a profound awareness of the English language, which L2 students have not developed. The language awareness of the L1 students – both tacit and explicit – is the result of life-long immersion in English, of linguistic and cultural in-dwelling, which L2 students do not have access to. Once again, when items tap into such language awareness, it raises issues as to what is being tested. Such items may measure more 'L1-ness' and less minimum competency in writing and reading. Thus it seems that these L2 students needed to have sufficient language proficiency before they could fully demonstrate their literacy skills in reading and writing, although the results of some other studies (e.g., Cummins, 2000) may reject this view.

We also found that some students came from family backgrounds that facilitated their literacy development. We probed further, when one L1 student kept saying he had a sense of what 'sounded right.' He explained,

Well, for me it is because it's been driven into my brain day after day, because my parents will never let a grammatically flawed sentence slip by. They'll say, 'Yeah, change it, and say it again.' Something like that. Yeah. I just have ... just the instinct for what sounds right. (L1 S-N, S\_B)

Family background is a crucial factor in students' literacy development. The more parents provide their children with resources or feedback on learning at home, the better the chance that the children will achieve academic success in school (Bianchi & Robinson, 1997). Findings from the present study highlight the importance of parental influence as well as the positive relationship between parental educational level and students' literacy development reflected by their test performance on the OSSLT.

*Test-taking processes*

When we asked detailed questions about how students approached each of the reading and writing tasks on the OSSLT, we came to understand better what was in the students' minds when they tackled each test task. The following student accounts indicate how the students said they approached reading on the test. On the whole, L2 students seemed to be more strategic in processing the reading tasks in comparison with their L1 counterparts, who, in turn, were more systematic: 'I first looked at the questions. Then I knew what I was looking for. Then I looked back, and then I answered the questions' (L1 S-V, S\_B).

I think I read it and then I read the last five paragraphs over and went to the questions and then I answered. And whenever I didn't find the answer, I didn't know it, I just looked backward and so I didn't have to read the whole thing. (L2 S-L, S\_B)

What I did was I look at the question, look at all possible answers, go through and find it, and I just do that for all the questions, and then I go through and read the passages just to make sure that I didn't miss anything when I did the questions. So it's like a way of rechecking it, without having to re-read the same parts over and over and over again. (L1 S-St, S\_B).

Similar findings were found with the writing tasks. L2 students seemed unable to step back and systematically manage the writing tasks: 'I simply thought and wrote without taking any note' (L2 S-E, S\_B). Conversely, L1 students seemed to approach the task much more systemically:

So I had one point – let's say, 'You shouldn't eat junk food because it's bad for your health' – and then I would list all the reasons why it's bad for your health and what they can do to you and I would... little kind of notes. And then I would just take those and make them into sentences when I start writing my essay. (L1 S-L, S\_A)

We also found that some L2 students did not know how to provide direct evidence to support their ideas. One L2 student started writing about the negative consequences of eating junk food. From the flow of his paragraph, it seems that his way of thinking was indirect or spiral. In his first paragraph, he wrote, 'Junk food is not healthy. It contains a lot of fat. Research has indicated that one piece of

MacDonald hamburger contains about...’ However, after writing about junk food as unhealthy, he began his second paragraph by pointing out, ‘You don’t have to eat fast food all the time,’ and, ‘That is not necessary.’ The sentences are both related to junk food but provide no support for his initial discussion of the unhealthy features of junk food. In this case the OSSLT is eliciting information about gaps in these students’ understanding of effective writing in English. In other words, there is validation evidence here that the test measures what it intends to measure. If information on students’ test performance on such items were communicated back to teachers in schools, teachers would be able to target their instruction to address these gaps. In an ideal relationship between testing and classrooms, the test measures what it intends to measure and the information about performance is communicated back to the stakeholders who learn from the test and improve the quality of instruction.

Another related aspect mentioned in the students’ accounts is the role of test preparation in their actual processing of the test tasks. When we asked a student how he knew to use direct quotations in the news report writing task, he remarked, ‘Well, . . . we had a meeting before the literacy test with Mrs. X from Student Services . . . she told us that we need to have a quotation or some quotations from people involved’ (L1 S-N S\_B). We found that all students involved in this study benefited from test preparation organized by their teachers. They all received test preparation before they took the OSSLT.<sup>7</sup> In both schools, teachers downloaded the sample literacy tests from the Internet, so students had the chance to go through the test and become familiar with the test format. ‘Our teacher always says like we’ll learn [by taking sample tests] . . . you will be a better writer if you read more’ (L1 S-Sa, S\_B).

### *Test topics*

We did not find that students discussed their preferences in terms of test topics. Some students commented that they knew less about a particular subject, such as fast food or customer service. Some L2 students had never encountered the concept of customer service. What we found was that L1 and L2 students provided varying accounts of what they wrote about based on the issues that concerned them. For example, in one of the writing tasks, they were asked to write down the one thing they want to change and why. One of the L2 students mentioned that she wanted to change the visa law regarding immigrants. That was due to the difficulties she and her family/friends experienced.

Another L2 student from South Korea said he wanted to change the allocation of resources. He commented it was unfair that some countries in this world have so few resources (like his country) and other countries have so many resources (like the United States and Canada). One of the topics written about by L1 students was obesity. For example, one student wrote that he wanted to change Highway 401 into a road with bicycle lanes so people could exercise and become healthy.

These topic preferences reflect students' choice of culturally familiar topics in writing. It seems that cultural schemata affect students not only in processing and comprehension of reading texts (Pritchard, 1990), but also in providing them with ideas for writing tasks.

### Conclusion

This study was undertaken as the second phase of a three-phase study designed to investigate the differential impact of the OSSLT on L1 and L2 students. In phase one, differences in test scores on reading and writing tasks on the OSSLT were investigated in relation to group membership (L1 or L2). On the basis of those findings, the present phase of the study focused on the differential impact of the OSSLT test types and test formats on L1 and L2 test-takers by eliciting their own accounts of reading and writing tasks on the 2006 test. As a follow-up to the study reported here, in phase three, questionnaires were distributed to students in a number of Ontario secondary schools across Grade 10 (pre-test) and Grade 11 (post-test) groups.

We took a general and holistic approach to understanding the processes these students engaged in while writing the OSSLT, a large-scale provincial test in Ontario. Although the retrospective verbal accounts of the test by L1 and L2 participants used in this study were not elicited as evidence of these students' concurrent thinking processes during the test, this retrospective approach avoided intruding into the students' actual testing processes while allowing us to listen to their accounts of their test experiences immediately after they took the test. We identified repeated patterns of responses and related those responses to L1 and L2 group membership and OSSLT test constructs. This method provided information on students' perceptions of test content (topics), test task properties (test types), and test-taking processes (including strategies). In addition, such information informed us about test constructs evident in these students' accounts, which in turn helped us to better understand their test performance.

We recognize that we did not control, in the design of our research, for L2 students' English proficiency, although all of the participants were functioning in high school classes and were eligible to write the test. At times some of our participants may not have been able to account for their processes because they did not have the language resources to do so, and with only 16 students participating in the study, considerable caution must apply in interpreting the results or assuming they may apply to other students or situations.

As has been noted in the research literature, an individual's mind and thinking is influenced or shaped by cultural and social factors. Sociocultural influences include 'the sets of values, beliefs, experiences, communication patterns, teaching and learning styles, epistemologies inherent in the students' cultural backgrounds, and the socioeconomic conditions prevailing in their cultural groups' (Solano-Flores & Nelson-Barber, 2001, p. 555). Cultural knowledge plays an indispensable role in students' comprehension of reading (Garcia & Gopal, 2003; Ginther & Stevens, 1998; Steffensen & Joeg-Dev, 1984). Research has shown that if students possess test writers' schemata, they can understand what is stated and make inferences more effortlessly, compared with unfamiliar schemata from the test writer. Ortiz (2005) claimed that if test-takers include those who have not been raised in the culture from which the test samples, which is the situation of the OSSLT for L2 students, then test validity needs to be discussed with caution. Further, Salvia and Ysseldyke (1991) (as cited in Ortiz, 2005) commented that it might be improper to use the norms of a test to evaluate a student's current performance or predict her future performance if the test population contains a group whose general background experiences are distinctively different from the overall population on which the test is standardized, as is the case of the OSSLT in this study. Using protocol analysis is a useful means of providing evidence to test developers that their tests meet validity criteria when the test-taker population is culturally heterogeneous. One step further in understanding students' accounts of tests would be a discursive approach (Huhta, Kalaja, & Pitkanen-Huhta, 2006), which can offer more context-sensitive and situation-variant information about the construction of tests from the viewpoint of test-takers.

In sum, this study identified five recurring categories in L1 and L2 test takers' accounts of the reading and writing tasks on the OSSLT. First, there were differences across these groups in their accounts of test difficulty. These differences may not, however, be due entirely

to the reading and writing constructs, which the test was designed to measure. Rather, the tasks may be measuring constructs that are not of central interest to the OSSLT (e.g., language proficiency, knowledge of specific vocabulary items, such as invention), or culturally embedded concepts (e.g., customer service agents). Second, although some OSSLT test formats appear to distinguish L2 from L1 test performance (Cheng et al., 2007), variability in the accounts of test-takers within the L1 and L2 groups considered here suggested that the range of reading and writing tasks is appropriate to the purposes of the test. Third, the groups considered here differed considerably in their test-taking strategies. The L1 group reported the use of text-based inference and elimination as a recurring strategy, and the L2 group said they often resorted to 'guessing' or matching keywords in questions and texts. Fourth, the L1 groups often indicated that an answer just 'sounded right.' They reported that their tacit and profound awareness of the English language, which L2 students have not developed, allowed them to respond to a task or item on the test. When items tap into such language awareness, it raises issues as to what is being tested. Such items may measure more 'L1-ness' and less minimum competency in writing and reading at the Grade 10 level. Further, the role of parents, their valuing of education and their support of learning at home, was highlighted as a key factor in test performance, as was the important role that test preparation plays – particularly for L2 students who are unfamiliar with tests in general and Canadian tests in particular.

On the whole, L2 students seemed to rely more on compensatory coping strategies in processing OSSLT tasks, whereas their L1 counterparts tended to be more systematic. Such differences across the L1 and L2 groups considered here may undermine interpretation of test results and they need to be further investigated. Finally, the test did elicit a range of topics across and within the groups that appeared to tap into concerns of the individual students considered here. Examining differences in the accounts of the OSSLT across L1 and L2 test-taker groups provide validation evidence and also helps to clarify issues of concern. These issues can be addressed as part of ongoing test validation.

**Liying Cheng** is an associate professor and a director of the Assessment and Evaluation Group (AEG) at the Faculty of Education, Queen's University, Canada. Her primary research interests are the impact of large-scale testing on instruction, and the relationship between assessment and instruction in EFL/ESL classrooms.



**Janna Fox** is an associate professor and a director of the Language Assessment and Testing Research Unit within the School of Linguistics and Applied Language Studies of Carleton University, Canada. Her principal scholarly interests are in language curriculum, educational reform, and testing and assessment processes and practices in test development and validation.

**Ying Zheng** is a doctoral student at the Faculty of Education, Queen's University. Her research interests focus on the relationship between individual differences and second/foreign language test performance and quantitative research methodology in language testing.

### Notes

- 1 ESL and ELD students are the distinct categories of language learners defined by the Ontario curriculum (Ministry of Education and Training, 1999). Both ESL and ELD students are subsumed under the term second language (L2) students within the present research context.
- 2 Information on OSSLT test development and field-testing can be found at <http://www.eqao.com/Test/2007/OSSLT/EN/faq.html>. According to the EQAO, the test is overseen by committees of Ontario educators and assessment experts, such as an Item Writing Committee, composed of curriculum and literacy experts who create test questions (items) and often pilot test them in their classrooms; an Assessment Development Committee, which ensures that the test matches curriculum expectations; a Sensitivity Committee, which ensures that the test content and reading passages are not biased against any student groups; and a Psychometric Expert Panel, composed of national and international experts in the science of assessment, who ensure the validity of the test.
- 3 Each of the 16 participants described in this article took part in two focus groups (one on reading and one on writing). In fact, we had slightly different combinations of students in each focus group, depending on their availability and the time that was allowed for us to conduct the focus groups, over school lunch hours and/or before the end of the school day.
- 4 Our research team conducted the data analysis manually in discussions at each stage of the analysis. Every member of the research team was involved with the data collection and analysis. The interaction of the team and the analysis of the data produced the key categories/characteristics of students' test-taking processes.
- 5 Extracts from our data are identified as follows. The school participating on March 29, 2006 was named School A and the one on March 30, 2006 was named School B, which are shortened to S\_A and S\_B. When students' accounts of the test are quoted, students identified as L1 or L2 students, as

well as their initials. For example, L1 S-J means L1 Student Jack. All names are, however, pseudonyms.

- 6 Fox and Cheng (2007) reported on data collected from another pool of students from a different school board in a different city.
- 7 Such test preparation is not consistently provided to students taking the OSSLT across the province and so may affect L2 test-takers who are unfamiliar with test formats and test-taking strategies (see Fox & Cheng, 2007).

### Acknowledgement

This research was made possible by a standard research grant from the Social Sciences and Humanities Research Council of Canada. We acknowledge the generous support of the Education Quality and Accountability Office in Ontario for allowing us access to the 2006 Ontario Secondary School Literacy Test items before its administration. We thank all the students who took the time to participate in our study. Acknowledgement also goes to Hong Wang and Baktygul Ismailova for conducting the study in the two schools.

### References

- Abedi, J., Leon, S., & Mirocha, J. (2003). *Impact of students' language background on content-based assessment: Analyses of extant data* (CSE Technical Report No. 603). Los Angeles: University of California, National Centre for Research on Evaluation, Standards, and Students Testing.
- Atkinson, D. (2005). Situated qualitative research and second language writing. In P. Matsuda & T. Silva (Eds.), *Second language writing research: Perspectives on the process of knowledge construction* (pp. 49–64). Mahwah, NJ: Erlbaum.
- Banerjee, J. (2004). Qualitative analysis methods. In *Reference Supplement to the preliminary pilot version of the manual for relating language examinations to the common European framework of reference for language: Learning, teaching, assessment* Section D. Strasbourg: Council of Europe, Language Policy Division.
- Banerjee, J., & Luoma, S. (1997). Qualitative approaches to test validation. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education: Language testing and assessment: Vol. 7* (pp. 275–287). Dordrecht, The Netherlands: Kluwer.

- 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.
- Bracewell, R. (1994). Withered wisdom: A reply to Dobrin. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (pp. 290–294). Thousand Oaks, CA: Sage.
- Cavanaugh, J.C., & Perlmutter, M. (1982). Metamemory: A critical examination. *Child Development*, 53, 11–28.
- Cheng, L., Klinger, D., & Zheng, Y. (2007). The challenges of the Ontario Secondary School Literacy Test for second language students. *Language Testing*, 24(2), 185–208.
- Clapham, C. (1998). The effect of language proficiency and background knowledge on EAP students' reading comprehension. In A.J. Kunnan (Ed.), *Validation in language assessment: Selected papers from the 17th language testing and research colloquium* (pp. 141–168). Mahwah, NJ: Erlbaum.
- Cohen, A.D. (1984). On taking language tests: What the students report. *Language Testing*, 1, 70–81.
- Cohen, A.D. (1997). Toward enhancing verbal reports as a source of insights on test-taking strategies. In A. Huhta, V. Kohonen, L. Kurki-Suonio & S. Luoma (Eds.), *Current developments and alternatives in language assessment* (pp. 339–366). Kopijyva, Finland: University of Jyväskylä.
- Cohen, A.D. (2007). The coming of age for research on test-taking strategies. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 89–112). Ottawa, ON: University of Ottawa Press.
- Collier, V.P. (1987). Age and rate of acquisition of second language for academic purpose. *TESOL Quarterly*, 21(3), 617–641.
- Collier, V.P. (1988). The effect of age on acquisition of a second language for school. *New Focus*, 2. Retrieved June 22, 2007, from <http://www.ncele.gwu.edu/pubs/classics/focus/02aage.htm>.
- Collier, V. (1992). A synthesis of studies examining long-term language minority student data on academic achievement. *Bilingual Research Journal*, 16, 187–121.
- Cooper, M., & Holzman, M. (1983). Talking about protocols. *College Composition and Communication*, 34, 284–293.
- Crain-Thoreson, C., Lippman, M.Z., & McClendon-Magnuson, D. (1997). Windows on comprehension: Reading comprehension processes as revealed by two think-aloud procedures. *Journal of Educational Psychology*, 89(4), 579–591.

- Cumming, A. (1990). Metalinguistic and ideational thinking in second language composing. *Written Communication*, 7(4), 482–511.
- Cumming, A. (1996). Introduction: The concept of validation in language testing. In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 1–13). Clevedon, UK: Multilingual Matters.
- Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics*, 11(2), 132–149.
- Cummins, J. (1982). *Tests, achievement, and bilingual students*. Arlington, VA: National Clearinghouse for Bilingual Education.
- Cummins, J. (1983). Conceptual and linguistic foundation of language assessment. In S.S. Seidner (Ed.), *Issues of language assessment: Vol. 2. Language assessment and curriculum planning* (pp. 7–16). Wheaton, MD: National Clearinghouse for Bilingual Education.
- Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire*. Clevedon, UK: Multilingual Matters.
- DiCerbo, P. (2000). Lessons from research: What is the length of time it takes limited English proficient student to acquire English and succeed in an all-English classroom? *Issue & Brief*, 5. Retrieved June 22, 2007, from <http://www.nclae.gwu.edu/pubs/issuebriefs/ib5.htm>.
- Duran, R.P. (1989). Testing of linguistic minorities. In R.L. Linn (Ed.), *Educational measurement* (3rd ed). (pp. 573–587). New York: Macmillan.
- Education Quality and Accountability Office. (2002). *Ontario Secondary School Literacy Test, February 2002: Report of Provincial Results*. Retrieved July 23, 2007, from <http://www.eqao.com/results/results.aspx?grade=10&year=2001&Lang=E&submit=View+Results>.
- 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](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](http://www.eqao.com/pdf_e/04/04P002e.pdf).
- Education Quality and Accountability Office. (2006). *Be informed. Be ready*. Retrieved July 23, 2007, from [http://www.eqao.com/pdf\\_e/06/06P056e.pdf](http://www.eqao.com/pdf_e/06/06P056e.pdf).
- Ericsson, K.A., & Simon, H.A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Ericsson, K.A., & Simon, H.A. (1987). Verbal reports on thinking. In C. Faerch & G. Kasper (Eds.), *Introspection in second language research* (pp. 24–53). Clevedon, UK: Multilingual Matters.

- Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurement*, 27(3), 209–226.
- Firestone, W., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis*, 20(2), 95–113.
- Fox, J., & Cheng, L. (2007). Did we take the same test? Differing accounts of the Ontario Secondary School Literacy Test by first and second language test takers. *Assessment in Education: Principles, Policy & Practice*, 14(1), 9–26.
- Garcia, P., & Gopal, M. (2003). The relationship to achievement on California high school exit exam for language minority students. *NABE Journal of Research and Practice*, 2003 (Winter), 123–137.
- Garner, R. (1988). Verbal report data on cognitive and metacognitive strategies. In C.E. Weinstein & E.T. Goetz (Eds.), *Learning and study strategies: Issues in assessment, instruction, and evaluation* (pp. 63–76). London: Academic.
- Ginther, A., & Stevens, J. (1998). Language background, ethnicity, and the internal construct validity of the advanced placement Spanish language examination. In A.J. Kunnan (Ed.), *Validation in language assessment: Selected papers from the 17th language testing and research colloquium* (pp. 169–194). Mahwah, NJ: Erlbaum.
- Gosden, H. (1996). Verbal reports of Japanese novices' research writing practices in English. *Journal of Second Language Writing*, 5(2), 109–128.
- Hayes, J., & Flower, L. (1980). Identifying the organization of writing processes. In L.W. Gregg & E.R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Hillsdale, NJ: Erlbaum.
- Huhta, A., Kalaja, P., & Pitkanen-Huhta, A. (2006). Discursive construction of a high-stakes test: The many faces of a test-taker. *Language Testing*, 23, 326–350.
- Jennings, M., Fox, J., Graves, B., & Shohamy, E. (1999). The test-takers' choice: An investigation of the effect of topic on language test performance. *Language Testing*, 16(4), 426–456.
- Klinger, D., Cheng, L., & Zheng, Y. (under review). *Factors influencing ESL/ELD students' performance on the Ontario Secondary School Literacy Test*. Manuscript submitted for publication.
- Krashen, S. (1987). *Principles and practices in second language acquisition*. Englewood Cliffs: Prentis-Hall.
- Krashen, S.D. (2007, March). *Free voluntary Web-surfing*. Plenary speech delivered at the 9th International Conference and Workshop on TEFL and Applied Linguistics, Ming Chuan University, Taiwan.
- Krueger, R.A., & Casey, M.A. (2000). *Focus groups: A practical guide for applied research*. (3rd ed). Thousand Oaks, CA: Sage.

- Kunnan, A.J. (1998). Approaches to validation in language assessment. In A.J. Kunnan (Ed.), *Validation in language assessment* (pp. 1–15). Mahwah, NJ: Erlbaum.
- Kuusela, H., & Paul, P. (2000). A comparison of concurrent and retrospective verbal protocol analysis. *American Journal of Psychology*, 113(3), 387–404.
- Manchon, R., Murphy, L., & Roca de Larios, J. (2005). Using concurrent protocols to explore L2 writing processes; Methodological issues in the collection and analysis of data. In P. Matsuda & T. Silva (Eds.), *Second language writing research: Perspectives on the process of knowledge construction* (pp. 49–77). Mahwah: NJ. Erlbaum.
- Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.
- Ministry of Education and Training (1999). *English as a second language and English literacy development: The Ontario curriculum Grade 9 to 12*. Toronto, ON: Ministry of Education and Training, Canada.
- Nevo, N. (1989). Test-taking strategies on a multiple-choice test of reading comprehension. *Language Testing*, 6(2), 199–255.
- Nisbett, R., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259.
- Norris, S.P. (1991). Informal reasoning assessment: Using verbal reports of thinking to improve multiple-choice test validity. In J.F. Voss, D.N. Perkins & J.W. Segal (Eds.), *Informal reasoning and education* (pp. 451–471). Hillsdale, NJ: Erlbaum.
- O'Malley, J.M., & Chamot, A.U. (1990). *Learning strategies in second language acquisition*. New York: Cambridge University Press.
- Ortiz, S.O. (2005). Acculturational factors in psychological assessment. In R.L. Rhodes, S.H. Ochoa & S.O. Ortiz (Eds.), *Assessing culturally and linguistically diverse students: A practical guide* (pp. 124–136). New York: Guilford.
- Parry, K.J. (1987). Reading in a second culture. In J. Devine, P.L. Carrell & D.E. Eskey (Eds.), *Research in reading in English as a second language* (pp. 59–70). Washington, DC: Teachers of English to speakers of other languages.
- Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, 20(1), 26–56.
- Pritchard, R. (1990). The effects of cultural schemata on reading processing strategies. *Reading Research Quarterly*, 25, 273–295.
- Roessingh, H. (1999). Adjunct support for high school ESL. *TESL Canada Journal*, 17(1), 72–87.

- Roessingh, H., & Kover, P. (2003). Variability of ESL learners' acquisition of cognitive academic language proficiency: What can we learn from achievement measures? *TESL Canada Journal*, 21(1), 1–21.
- Ryan, K. (2002). Assessment validation in the context of high-stakes assessment. *Educational Measurement: Issues and Practice*, 21(1), 7–15.
- Sasaki, M. (2000). Effects of cultural schemata on students' test-taking processes for cloze tests: A multiple data source approach. *Language Testing*, 17(1), 85–114.
- Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics* (pp. 143–175). Cambridge, UK: Cambridge University Press.
- Selinger, L., & Shohamy, E. (1989). *Second language research methods*. Oxford, UK: Oxford University Press.
- Sepassi, F. (2003). How do learners of different language ability perform on the cloze? A verbal protocol analysis of EFL test takers performance on cloze tests. *Indian Journal of Applied Linguistics*, 29(2), 5–33.
- Smagorinsky, P. (1991). The reliability and validity of protocol analysis. *Written Communication*, 6, 463–479.
- Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching*, 38(5), 553–573.
- Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practices paradigms in the testing of English-language learners. *Educational Researcher*, 32(2), 3–13.
- Steffensen, M.S., & Joeg-Dev, C. (1984). Cultural knowledge and reading. In J.C. Alderson & A.H. Urquhart (Eds.), *Reading in a foreign language*. New York: Longman.
- Storey, P. (1997). Examining the test-taking process: A cognitive perspective on the discourse cloze test. *Language Testing*, 14(2), 214–231.
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
- Watt, D., & Roessingh, H. (2001). The dynamics of ESL drop-out: Plus ça change . . . *Canadian Modern Language Review*, 58(2), 203–222.
- Wijgh, I.F. (1996). A communicative test in analysis: Strategies in reading authentic texts. In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 154–165). Clevedon, UK: Multilingual Matters.
- Wu, Y.A. (1998). What do tests of listening comprehension test? A retrospection study of EFL test-takers performing a multiple-choice task. *Language Testing*, 15(1), 21–44.



## Appendix 1

### *Procedures for conducting focus group interviews*

#### General questions

- |  |     |    |
|--|-----|----|
| 1 Gender   | F   | M  |
| 2 Were you born in Canada?   | Yes | No |
| 3 Do you speak another language at home?   | Yes | No |
| 4 If yes, which language do you speak? _____   |     |    |
| 5 Is this your first time taking the test?   | Yes | No |
| 6 If no, when was your last time taking the test? _____  |     |    |
| 7 How do you like the OSSLT? How do you think you did on the test?<br>Do you think you have passed the test? _____ |     |    |

#### Focus group interview questions on the *writing* tasks

- 1 How did you like the writing tasks? Which one of the writing tasks – long (Section E: News Reports, and Section H: Opinion) *or* short tasks (Section D: Invention and Section G: Change) did you find easy or difficult? Why? How did you like the multiple-choice writing tasks in Section B and I?
- 2 Long writing task (Section E: News Report) – What did you see/get from the picture? How did you find the connection between the picture and the title? What were some of the things you wrote down on this topic? How did you address *who*, *what*, *where*, *when*, *why* and/or *how*? What was the length of your writing? Who did you have in mind when you wrote this news report?
- 3 Short writing task (Section G: Change) – What was the one thing you want to change? Why?
- 4 Multiple-choice writing tasks (Section I) – Which options did you choose (on each item)? Why? What is (or where is in the questions/choices) your evidence of your answer?
- 5 Have you paid attention to the grammatical aspects of the language (convention i.e., punctuation and capitalization, usage)? What kind of convention did you focus on?

#### Focus group interview questions on the *reading* tasks

- 1 How did you like the reading passages? Which one of the reading sections – Section A, C, F, J, K, or M did you find easy or difficult? Why?

2 Section F (Narrative Reading) – Ask all questions in this section. Ask ‘what,’ ‘why,’ ‘where’ questions.

- What answer did you choose in each question?
- Why did you choose that specific question?
- Where did you find evidence to support your choice?

If the choice is wrong, continue asking

- what do you think might have caused you to make that wrong choice (what was the problem/difficulty you encountered)?  
What did you have in your mind or you felt this was the right choice?

3 Section K (Dialogue narrative Reading)

Start from question 9, 10 (open response questions)

- Ask them what they answered, why they answered in that way, and their evidence.

If time permits, continue with the questions in sequence below. They are listed according to a combination of format, strategies/difficulty levels. The eliciting questions follow the same format in section F.

Q2, Q3, Q4, Q7, Q8/Q1, Q5, Q6