COMPUTER FAMILIARITY AND TEST PERFORMANCE ON A COMPUTER-BASED CLOZE ESL READING ASSESSMENT

by **Dennis Murphy Odo**
Georgia State University,
Atlanta, Georgia, USA
d.modo @ alumni.ubc.ca

Abstract

Researchers have raised questions regarding the connection between learner familiarity with computers and performance on computerized tests virtually since interest arose in studying the applicability of computers for assessment purposes. However, despite this longstanding attention, at present, there has been a surprising lack of research that explores the connection between computer familiarity and performance on computerized tests that fall outside of the traditional multiple-choice discrete-point tests that have historically predominated in the fielf of testing and assessment.

The current study aims to address this gap in previous research by examining the relationship between computer familiarity and computer-based test performance on a computer-based test of second language reading that is integrative rather than discrete-point. The study investigated the online reading ability of ESL students from one secondary school in a large city in western Canada (61 females and 59 males in the sample, ages 13-19, M=15.73). The students responded to a questionnaire about their computer familiarity and then completed an online multiple-choice cloze test. Contrary to other most other findings based on discrete-point tests, the results revealed that the familiarity variables do account for a small but significant amount of the variability in the computer-based test scores.

1. Introduction

North American schools have seen a dramatic increase in the number of English language learners (ELL). In the US, according to the National Clearinghouse for English Language Acquisition (2006), in 1990, out of every 20 public school students in grades K-12 only one was an English language learner (ELL). This proportion has increased to 1 in 9 at present and given current trends could be 1 in 4 in the next 20 years. A large number of Canadian urban centres also contain a substantial proportion of ELLs within their schools. If fact, approximately 47% of Toronto District School Board secondary students and 41% of elementary students have a language other than English as their first language (Toronto District School Board, 2012).

Thirty-seven percent of learners in Greater Vancouver districts speak a first language other than English (Garnett & Aman, cited in Canadian Counsel on Learning, 2008). These statistics clearly indicate that ELLs comprise a substantial proportion of our urban school population. Therefore, we must realize that meeting the needs of ELLs ensures that our schools are functioning as intended. However, the disturbing disappearance rates of ELLs from schools across the US (Hoffman & Sable, 2006; Kindler, 2002) and Canada (Radwanksi, 1987; Gunderson, 2009; Watt & Roessingh, 2001) raise questions about our effectiveness in supporting them in their learning. Menken (2008) pointed out that informative assessments are crucial to ensure that immigrant students' learning needs are met.

Personnel in school districts across the Lower Mainland region of British Columbia, Canada, as studied in this investigation employ a variety of diverse measures to assess reading ability. This practice presents a problem when their highly-mobile ELL student population moves between districts because there is very little meaningful assessment information that can be shared among schools across districts (Gunderson, Murphy Odo, D'Silva, 2010). An ESL Assessment Consortium (http://www.eslassess.ca) was formed to address this problem as well as a variety of related assessment issues. Members of the Consortium developed a standardized assessment measure for ESL and native speaking students that provides an approximate indicator of reading proficiency (Gunderson, Murphy Odo, & D'Silva, 2010). The main advantage of this measure is that it provides all of the districts participating in its development with a common, locally-normed reading assessment that allows them to share information about reading proficiency when ELL students move between districts. The assessment has been administered to thousands of students in 12 different school districts since 2009.

The assessment, known as the Lower Mainland English Reading Assessment (LOMERA), has served districts, teachers, and students well. This type of assessment (i.e., multiple-choice cloze) has also been validated by extensive research as an accurate indicator of reading proficiency (e.g., Espin & Foegen, 1996; Hale et al., 1989; Oller & Jonz, 1994). Consortium members decided that the development of an online version of LOMERA was needed and that the issue of comparability between the paper-and-pencil version and the online version should be explored and whether test takers' familiarity with computers would have an impact on their computerized LOMERA performance. A careful review of relevant research literature did not provide a satisfactory answer to these questions for an assessment tool like

LOMERA so further research seemed necessary. The present study was designed to explore these questions.

2. Background to the study

Second language assessments often wield considerable power over the educational trajectories of students' lives (Shohamy, 2000). This is troubling because these tests can systematically discriminate against particular groups as a result of test bias in methods and materials (Menken, 2008). At present, the use of computerized forms of assessment is rapidly expanding to the point where we may need to see appropriate use of technology as a facet of language ability. Chapelle and Douglas (2006) suggest that "communicative language ability needs to be conceived in view of the joint role that language and technology play in the process of communication" (p. 108). Thus, we need to expand our view of language ability to see it as "the ability to select and deploy appropriate language through the technologies that are appropriate for a situation" (p. 107). Given these accelerating developments in language assessment technology, there is an urgent need for research that ensures development of computerized assessment tools and yields results that are valid, reliable and fair for all test takers. In particular, test designers and users need to ensure that test-taker computer familiarity is not inordinately affecting his or her test performance.

From the outset of research into this area computer familiarity has long been a background variable hypothesized to affect test takers' performance on computerized language tests. Kirsch, Jamieson, Taylor and Eignor (1998) conducted a study for Educational Testing Service (ETS) as part of the development of the computer-based version of the Test of English as a Foreign Language (TOEFL). They surveyed a sample of almost 90,000 TOEFL examinees from a wide variety of first languages. Their findings revealed small differences in computer familiarity based on age, gender and reason for taking the test. They reported larger discrepancies in computer familiarity for examinees from Japan and African countries. They also observed a small but significant difference in the performance on the paper-based test scores but, surprisingly, their study did not investigate the relationship between familiarity and computer-based test performance.

Results of the majority of recent studies demonstrate that computer familiarity does not affect performance on a computer-based language test. An investigation of 1,200 TOEFL

examinees from a wide variety of first language groups determined that when English language ability was taken into account there was "no meaningful relationship between level of computer familiarity and level of performance on the CBT [i.e. computer based test] language tasks" (Taylor, Kirsch, Eignor, & Jamieson, 1999, p. 265). These researchers concluded that there was no "adverse relationship between computer familiarity and computer-based TOEFL test performance due to lack of prior computer experience" (Taylor et al., 1999, p. 219). Similar results were reported with Saudi medical students who were taking a locally-developed EFL reading comprehension exam (Al-Amri, 2008). Sawaki (2001) conducted an extensive review of research literature in educational and psychological measurement as well as in ergonomics, education, psychology, and L1 reading research. Her review into the cross-mode effects for second language reading examinees similarly determined that "computer familiarity…does not seem to manifest itself in test scores" (p. 44).

Other studies with elementary school students that included both native speakers and second language learners concluded that reading comprehension test scores were not affected by computer literacy, but the authors did note that students with weaker computer skills were particularly disadvantaged when they had to scroll through text (Higgins et al., 2005). Some researchers who did find differences in performance across testing modes also noted that computer familiarity was not related to this performance difference (Clarania & Wallace, 2002).

In contrast to the cross-mode comparability research that indicates computer familiarity is not a significant factor, there is some other research which seems to indicate that it could depend on the devices being used. One potential familiarity-related cause in discrepant cross-mode test performances mentioned was the affordances of paper-based tests that are not readily available in a computerized mode. For instance, test takers may be more comfortable with being able to highlight or take notes on the printed page, but not on the computer screen. It is argued that their inability to interact with the computer test as they do with the paper may prevent them from performing as well as they could. Some authors conjecture that computer programs can probably be designed that will allow for more interactivity with the text (Choi et al., 2003). With the advent of devices such as tablet PCs, this is certainly the case.

Another computer affordance-related area of inquiry was examinees' ability to manipulate the accessories of the computer. This specific area of inquiry related to test interactivity and computer familiarity that was mentioned in at least two studies. Pomplun, Frey

and Becker (2002) compared paper and computer versions of a reading placement test for secondary and post-secondary students. Their data revealed that, in a number of cases, differences in respondents' performance across the two modes seemed to be caused by the differences in their response speed associated with the use of a mouse compared to a pencil. That is, those who performed better on the computerized test tended to be more adept at using the mouse. Other studies into secondary students' ability to answer open-ended (i.e., short answer and essay) questions indicated that learners who have more experience with keyboarding performed better on the computer-based test (Russell, 1999). These findings imply that facility with using computer devices may have some bearing on CBT performance.

This discrepancy in research findings appears to warrant further investigation. In particular, the bulk of research into computer familiarity appears to have been conducted with traditional discrete-point tests. The present study has explored the influence of familiarity on test performance with a form of language assessment that as yet does not appear to have been investigated (i.e., multiple-choice cloze). This investigation should help broaden the research base on computer familiarity by addressing how this variable interacts with performance on various types of computer test tasks. The main objective of this study was to identify whether learners' computer familiarity predicted their performance on the computerized LOMERA.

3. Methodology

3.1. Selection and recruitment of participants

The research site was a large secondary school in a major city on the west coast of Canada. The school serves learners from grade 8 through 12. A sizable proportion of the student body is comprised of recent immigrants to Canada. The school has an ESL program that serves hundreds of students of all ages, grades and ESL levels. Participants in this study came from a variety of different countries including Brazil, China, Iran, Korea, Mexico, Philippines, Russia, and Vietnam. They were recruited by this author with the assistance of their ESL teachers. The number of participants that was recruited for the study was 60 for each testing group which totalled 120.

The primary purpose for selecting these students is that secondary-school-aged participants have often been overlooked in the second-language assessment research literature

(Snow, 2008). These participants were also chosen because they are the most suitable group to help establish whether the paper-based and computer-based versions of LOMERA are equivalent in their ability to discriminate ESL learners' reading proficiency. This is because the school is from the region where the test is normed and the study participants are in the group for whom the test was designed – secondary school ESL students.

3.2. Instruments

The questionnaire used in this study was paper-based to avoid any possible confounds caused by presenting it through different media. Information contained in these questionnaires was taken from past research studies on computer familiarity (Taylor et al., 1999). The questionnaire included eighteen questions about test takers' computer familiarity. The familiarity questionnaire asked participants to indicate how often computers are available to them in various locations such as home and school. They are presented with four categorical options from which to choose labelled "once a week or more often," "between once a week and once a month," "less than once a month," and "never." They are then asked about their comfort level with various aspects of computer use such as taking tests on a computer. Their response is divided into four categories which are "very comfortable," "comfortable," "somewhat comfortable," and "not comfortable." Finally, they are asked about how often they use a computer for various tasks and they are to choose among four categories labelled "more than once a day to once a week," "less than once a week to once a month," "less than once a month," and "never."

The questions were scored by indicating which category was selected and the results were entered into SPSS. The questionnaire was informally piloted with several second-language speakers beforehand to ensure that the language of the questions was comprehensible to participants.

3.3. Design and procedure

A modified counterbalanced design (see Chihara et al., 1992) was selected for this investigation for two reasons. First, this design allowed for comparison of participants' performance on the paper-based LOMERA test with their performance on the computer-based versions of the same test. This is essential because the central goal of this investigation was to determine whether there is a statistically significant difference in test performance between these two testing modes.

Second, there is no control group in the traditional sense because each of the randomly-assigned groups will be serving as the control for the other group. Nevertheless, the paper version of the test could be considered to be a type of control while the computer could be thought of as the treatment condition because the paper test had already been validated and its reliability had been confirmed. Gribbons and Herman (1997) explain that counterbalanced designs allow for all groups to take part in more than one randomly-ordered treatment and control conditions. As is illustrated in Figure 1 below, at the outset of this study half of the participants were randomly assigned to take the paper-based version of LOMERA and the other half took the computer-based version. One month later test takers switched roles and took the test in the mode that they had not initially taken it.

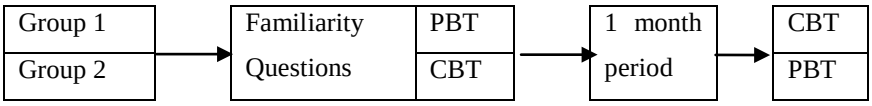

Figure 1. Diagram representing the procedures in the study.

Data collection began after the present author contacted teachers and district administrators and explained the study to obtain permission to conduct research at the school. An ethics review for the institutional Behaviorial Research Ethics Board (BREB) was completed and authorization was gained to study human subjects. Participants were assured that their identities would remain confidential during and after the research. Informed consent for students to participate in this research was then obtained from students and their parents. When all permissions were granted, schools were visited, participants were randomly assigned into either a computer-first or a paper-first group, and data collection began.

One hundred and twenty participants were administered the paper-based and computer-based versions of the LOMERA test. Half of the sample was given the paper version and the other half was given the computer version. Approximately four weeks later the group that had previously taken the paper-based test took the computerized version and vice versa. Although there is no prescribed minimum period to wait between test administrations (Kauffman, 2003), the interval chosen for the second administration of the test for this study was four weeks. A four-week interval should be an adequate minimum period to wait before administering the second test for two reasons discussed by Kauffman (2003). First, if the interval is too short (e.g.

a few days), learners will be more likely to remember answers to specific questions and test taking strategies learned through completing the first test. These advantages will then inordinately enhance their subsequent test performance (i.e. practice effect). Similarly, if the period between test administrations is too long, it becomes increasingly likely that other extraneous variables like maturation or instruction will affect performance. A four-week interval was thought to allow enough time for learners not to recall the test but without giving too much time to introduce other potential threats to validity.

In addition to taking both versions of the test, participants also completed a written questionnaire that asked them to report on their familiarity with computers. Results from the familiarity surveys were then compared with computer test performance to determine the relationship between computer familiarity and computer test performance.

4. Results and discussion

Upon determining the degree of comparability between the two measures, participants' results from the computer familiarity questionnaire were compared with their performance on the computer-based testing mode using a multiple regression analysis. First, the categorical variables elicited by the questionnaire were recoded into dummy variables to allow for meaningful interpretation of the results. The categories in these new recoded variables were then compared with a constant to establish whether self-reported comfort with computers, comfort with computer tests and number of tests taken on computer (which were taken as indicators of computer familiarity) predicted differences in test takers' computer scores.

The purpose of this analysis was to establish the degree to which computer familiarity predicted their performance on the computer version of LOMERA. Therefore, the variables' ability to predict performance on the computer version of LOMERA is viewed as an indicator of the effect of computer familiarity on computer test performance. Results of these analyses are reported in the section below.

4.1. Demographic and descriptive findings

The sample size used for this study was comprised of 120 individuals. There were 61 females and 59 males in the sample. The mean age was 15.73 (SD = 1.67). The youngest participant was thirteen and the oldest was twenty. The following tables (1, 2, and 3) contain breakdowns for

number and percentage of students in each grade, each ESL level and each first-language group designation.

Table 1. Frequencies of school grade

Grade		Frequency	Percent	Valid Percent
Valid	8	17	14.2	14.3
	9	17	14.2	14.3
	10	27	22.5	22.7
	11	33	27.5	27.7
	12	25	20.8	21.0
	Total	119	99.2	100.0
Missing	99	1	.8	
Total		120	100.0	

Table 1 above shows that there are a slightly larger proportion of participants in the higher grades than in the lower grades.

Table 2. Frequencies of ESL level

	ESL Level	Frequency	Percent	Valid Percent
Valid	1	29	24.2	24.2
	2	31	25.8	25.8
	3	38	31.7	31.7
	4	22	18.3	18.3
	Total	120	100.0	100.0

Table 3 demonstrates that there is a generally comparable proportion of levels although there are marginally fewer test takers in level four.

Table 3. Frequencies of first language group

		Frequency	Percent	Valid Percent
Valid	Mandarin	12	10.0	11.4
	Vietnamese	8	6.7	7.6
	Tagalog	22	18.3	21.0
	Kinyarwanda	1	.8	1.0
	Ilocano	8	6.7	7.6
	Ilongo	2	1.7	1.9

	Bisaya	5	4.2	4.8
	Cantonese	5	4.2	4.8
	Chinese	24	20.0	22.9
	Spanish	2	1.7	1.9
	Korean	1	.8	1.0
	Russian	2	1.7	1.9
	Arabic	1	.8	1.0
	Portuguese	1	.8	1.0
	Bahnar	2	1.7	1.9
	Tamil	1	.8	1.0
	Tigrinya	2	1.7	1.9
	Burmese	1	.8	1.0
	Jarai	4	3.3	3.8
	Karen	1	.8	1.0
	Total	105	87.5	100.0
Missing	99	15	12.5	
Total		120	100.0	

A number of test takers (12.5%, N = 15) did not report their first language and among those who did a number identified their first language as "Chinese" (20%, N = 24), but did not specify whether or not it was Mandarin, Cantonese or some other dialect of the language. It should also be noted that a sizable proportion of the study participants were native speakers of Asian languages. Although immigrants from Asia make up the largest group moving to urban centers around North America (Statistics Canada, 2010), this sample may not be entirely representative of immigrant demographics in Canada.

4.2. Results of Computer Familiarity Questionnaire analysis

Table 4. Descriptive statistics for mean scores and Standard Deviations for familiarity dummy variables

	Mean	SD	N
comptotal	63.23	18.566	120
somewhat comfort	.050	.218	120
comfortable	.366	.483	120
very comfort	.550	.499	120
somewhat comfort	.266	.444	120
comfortable	.516	.501	120
very comfort	.158	.366	120
1-2	.383	.488	120
3-4	.250	.434	120
5 or more	.141	.350	120

Descriptive statistics for the criterion variable, computer test total score, and each predictor variable used in the regression analysis are reported in Table 4. Figures included in the table are the mean score for the variable, standard deviation and number of participants.

Table 5. Correlations between LOMERA computer test score and dummy variables representing self-reported comfort with computers, taking tests on computers and number of tests taken on computers

		Somewhat Comfort (computer)	Comfortable	Very Comfort	Somewhat Comfort (Test)	Comfortable	Very Comfort	1-2 Tests	3-4	5 or More
Pearson Correlation	Comptotal	-.019	-.096	.235*	-.092	.253*	-.103	.168*	.052	-.261*
	Somewhat Comfort (comp)	1.000	-.175	-.254*	.294**	-.237*	.005	-.024	-.044	-.093
	Comfortable		1.000	-.841**	.128	.009	-.235*	.076	.000	-.111
	Very Comfort			1.000	-.326**	.164*	.209*	-.045	.019	.079
	Somewhat Comfort (test)				1.000	-.623**	-.262*	.145	-.131	-.083
	Comfortable					1.000	-.448**	.077	.096	-.133
	Very Comfort						1.000	-.201*	.013	.217
	1-2 Tests							1.000	-.455**	-.320*
	3-4								1.000	-.235**
	5 or More									1.000

* Correlation significant at the .05 level
** Correlation significant at the .001 level

Correlations between the predictor variable "comptotal" or total LOMERA score on the computer and each of the dummy independent variables are found in Table 5. Test takers' self-reported data about how comfortable they feel using computers and how comfortable they feel taking tests on a computer were correlated with their overall scores on the computer version of LOMERA. The correlations among all of the predictor variables are all less than .3, which shows there is lack of a strong relationship among any of these variables. These low correlations also indicate that there is not a great deal of colinearity among the variables, which is an assumption that must be met to justify their use in the regression analysis.

Table 6. Summary of hierarchical regression analysis for variables predicting computer-based LOMERA score

Source	Sum of Squares	df	Mean Square	Adjusted R^2	F Value	Sig.
Regression	10905.20	9	1211.68	.206	4.42	.001
Residual	30115.72	110	273.77			
Total	41020.92	119				
Predictor Variable	B	SE B	β	t	p	
(Constant)	15.02	11.56		1.299	.197	
somewhat comfort (comp)	31.48	11.01	.371	2.859	.005	
comfortable	30.25	9.08	.789	3.329	.001	
very comfort	37.15	9.03	1.000	4.115	.000	
somewhat comfort (test)	13.79	7.20	.330	1.913	.058	
comfortable	16.06	6.73	.434	2.386	.019	
very comfort	10.22	7.47	.202	1.367	.174	
1-2 tests	4.35	4.10	.115	1.063	.290	
3-4	3.03	4.45	.071	.682	.496	
5 or more	-6.54	5.35	-.123	-1.221	.225	

a Predictors: (Constant), 5 or more, very comfort, comfortable, 3-4, somewhat comfort, very comfort, 1-2, somewhat comfort, comfortable
b Dependent Variable: comptotal

As displayed in Table 6 above, adjusted R square = .20; F (9, 110) = 4.42, $p < 0.01$. (Several dummy variables were not significant predictors in this model). The standard error of estimate is 11.56. A multiple regression analysis was conducted with the dummy variables created for each category from the questions that asked test takers about their "comfort with computers," "comfort with taking tests on computers" and "number of tests taken on computer." This type of analysis was chosen to investigate the degree to which variables associated with computer familiarity influence computer test performance. The null hypothesis was rejected based on evidence that

these variables associated with computer familiarity were statistically significant predictors of computerized LOMERA test performance (F (9, 110) = 4.42, $p < 0.01$). That is, there is some relationship between computer familiarity as it is measured in this study and computerized LOMERA test performance. As can be seen in Table 6 above, the combination of these predictor variables accounts for just over 20% of the variance in online LOMERA scores. Therefore, one conclusion is that while there is a statistically significant relationship between computer familiarity and computer test performance, familiarity did not exert an inordinate influence on the variability of computer test performance in this study. That is, while "comfort with computers" and "comfort with taking tests on computers" play some role in LOMERA computer test performance, it is not a predominant role.

The results of the present study establish that while the variables that relate to computer familiarity do have some impact on online LOMERA performance, they are not the dominant factors in determining computer test scores. These findings indicate that further research should be conducted to ascertain more specifically if there are any other as-yet unidentified factors related to computer familiarity that may be impacting computer test performance and what those factors might be. Additionally, other variables that are less obviously related to computer familiarity should be explored in future research as well.

4.3. Discussion

The findings of this study are generally consistent with other research literature on computer familiarity (see Clarania & Wallace, 2002; Higgins et al., 2005). Taylor et al.'s (1999) study of 1,200 TOEFL examinees reported that computer familiarity does not affect performance on a computer-based language test. They concluded that there was "no meaningful relationship between level of computer familiarity and level of performance on the CBT [i.e. computer based test] language tasks" (p. 265). Another investigation with Saudi learners taking a reading test reached similar conclusions (Al-Amri, 2008). Sawaki (2001) offered similar observations in an exhaustive review of computer assessment-related research literature from a wide variety of academic disciplines.

The only cautions mentioned in other studies was the difficulty some students had with scrolling down the screen (Higgins et al., 2005) or having to use other interactive computer hardware such as the mouse (Pomplun et al., 2002) and the keyboard (Russell, 1999). In the

present study, students only had to use the mouse for the test and they did not have to type anything using the keyboard. Nevertheless, the possibility exists that some test takers may have struggled slightly with the mouse. It might be advisable to investigate this aspect of the LOMERA test-taking experience more closely in the future.

Another noteworthy limitation of the online test taking experience mentioned by Choi et al. (2003) was examinees' inability to interact with the computer screen by "marking it up" as they might do with a paper test. During the testing sessions a colleague observed that this interactivity was not as constrained in the computer mode as it first appears. In fact, there were several examinee-improvised methods of interacting with the text of the online LOMERA that appeared to be advantageous for those who were able to discover them. Two examples of such inventiveness were the manipulation of font sizes to make the test easier to read and using the cursor to "highlight" portions of text upon which the test taker was focusing. Future research might explore how test takers invent alternative means of making the online test taking experience more interactive. Designers might then attempt to incorporate these adaptations into future test iterations.

5. Conclusions

The research question asked in this investigation was: Do L2 learners who are more familiar with computers achieve higher scores on a computer-based multiple-choice cloze reading assessment than those who are less familiar with computers?

Findings are generally consistent with other research literature that computer familiarity does not have an inordinate impact on a computer-based language test performance. Although there is a statistically significant relationship between indicators of computer familiarity and online LOMERA performance, these variables actually do not explain a great deal of the variance in computer-based LOMERA test performance. A corollary of this conclusion is that computer familiarity may not be an important consideration in deciding whether to adopt the online LOMERA in local districts. Nevertheless, future studies might explore whether there may be a threshold of familiarity after which it becomes less of an issue. It could be that most of those who took the test in the present study were beyond that threshold.

The findings for the computer familiarity component of the study have several implications for assessment practices with LOMERA in the lower mainland. As mentioned

above, local test administrators do not have to be unduly concerned about the computer familiarity of those taking the test. In fact, one anecdote serves as a reminder that learners can often surpass our expectations and teachers may sometimes underestimate the computer familiarity of their ESL students. At one point during the test administration the test was given to a group of lower-level students. The teacher was initially sceptical about whether several students in the group would be able to take the computer test because they were later-to-literacy learners who had not had a great deal of experience with computers. The teacher was surprised during the administration of the test when all of the students were able to navigate the test with much less difficulty than had been anticipated. This example illustrates that it may be worthwhile to allow the student to attempt the test before deciding that he or she would probably be better served by doing it on paper. Nevertheless, if teachers suspect that a test taker has had insufficient previous exposure to computers (e.g. using the test at a reception centre) they might consider developing some type of brief pre-screening instrument or protocol to ensure that they have the requisite computer skills to guarantee a valid and comparable test administration.

References

Al-Amri, S. (2008). Computer-based testing vs. paper-based testing: A comprehensive approach to examining the comparability of testing modes. *Essex Graduate Student Papers in Language & Linguistics*, *10*, 22-44.

Canadian Counsel on Learning (2008). Lessons in learning: Understanding the academic trajectories of ESL students. Accessed March 24, 2012. http://www.ccl-cca.ca/pdfs/LessonsInLearning/Oct-02-08-Understanding-the-acedemic.pdf.

Chapelle, C. A. (2008). Utilizing technology in language assessment. In E. Shohamy, & N. H. Hornberger (Eds.), *Encyclopedia of language and education*, 2nd Edition, Volume 7*: Language testing and assessment* (pp. 123–134). New York: Springer Publishers.

Chapelle, C. A. & Douglas, D. (2006). *Assessing Language to Computer Technology*. Cambridge: Cambridge University Press.

Chihara, T., Oller, J. W., Weaver, K. A.., & Chavez-Oller, M. A. (1992). Are cloze items sensitive to constraints across sentences? *Language Learning*, 27, 63-73.

Choi, I. C., Kim, K. S., & Boo, J. (2003). Comparability of a paper-based language test and a computer-based language test. *Language Testing*, *20*, 295-320.

Clariana, R. & Wallace, P. (2002). Paper-based versus computer-based assessment: Key factors associated with the test mode effect. *British Journal of Educational Technology*, 33, 593-602.

Dooey, P. (2008). Language testing and technology: Problems of transition to a new era. *ReCALL*, *20*, 21-34.

Espin, C. A. & Froegen, A. (1996). Validity of general outcome measures for predicting secondary students' performance on content-area tasks. *Exceptional Children*, *62*, 497-514.

Fraenkel, J. R. & Wallen, N. E. (2005). *How to Design and Evaluate Research in Education.* New York: McGraw-Hill.

Garnett, B. & Aman, C. (April 4, 2008). Academic performance of immigrant youth in Vancouver, Montreal and Toronto: The Vancouver context. Paper presented at 10th National Metropolis Conference in Halifax, Nova Scotia.

Gribbons, B. & Herman, J. (1997). True and quasi-experimental designs. *Practical Assessment, Research & Evaluation, 5(14)*. Retrieved August 9, 2010 from http://PAREonline.net/getvn.asp?v=5&n=14.

Gunderson, L. (2009). *ESL (ELL) Literacy Instruction: A Guidebook to Theory and Practice* (2nd ed.). New York: Routledge.

Gunderson, L., D'Silva, R. & Murphy Odo, D. (2010). *The Lower Mainland English Reading Assessment (LOMERA) Manual.* Vancouver: The Lower Mainland ESL Assessment Consortium. Retrieved August 9, 2011 from http://www.eslassess.ca/esl/.

Hale, G. A., Stansfield, C. W., Rock, D. A., Hicks, M. M., Butler, F. A. & Oller, J. W. (1989). The relation of multiple-choice cloze items to the Test of English as a Foreign Language. *Language Testing, 6*, 47-75.

Higgins, J., Russell, M., & Hoffmann, T. (2005). Examining the effect of computer-based passage presentation on reading test performance. *Journal of Technology, Learning, and Assessment, 3*. Retrieved June 9, 2010, from http://escholarship.bc.edu/jtla

Hoffman, L., & Sable, J . (2006). *Public Elementary and Secondary Students, Staff, Schools, and School Districts: School Year 2003-2004*. Washington, DC: National Center for Education Statistics

Kauffman, A. S. (2003). Practice effects. *Speech and Language Forum.* http://www.speechandlanguage.com/cafe/13.asp.

Kindler, A . (2002). *Survey of the States' Limited English Proficient Students and Available Education Programs and Services: 2000-2001 Summary Report.* Washington, D C: National Clearinghouse for English Language Acquisition

Kirsch, I., Jamieson, J., Taylor, C. & Eignor, D. (1998). *Computer Familiarity among TOEFL Examinees.* TOEFL Research report 59, March, 1998, Princeton, NJ: Educational Testing Service.

Menken, K. (2008). *English Learners Left Behind: Standardized Testing as Language Policy.* Toronto: Multilingual Matters.

National Clearinghouse for English Language Acquisition (2006). *How Has the English Language Learner Population Changed in Recent Years?* Washington, DC: NCELA; Accessed March 24, 2012. www.ncela.gwu.edu/expert/faq/08leps.html.

Oller, J. W. & Jonz, J. (Eds.).(1994). *Cloze and Coherence.* Cranbury, NJ: Bucknes University Press.

Pomplun, M., Frey, S. & Becker, D. (2002). The score equivalence of paper-and-pencil and computerized versions of a speeded test of reading comprehension. *Educational and Psychological Measurement, 62*, 337-354.

Radwanski, G. (1987). *Ontario Study of the Relevance of Education and the Issue of Dropouts.* Toronto: Ontario Ministry of Education.

Russell, M. (1999). *Testing on Computers: A Follow-Up Study Comparing Performance on Computer and on Paper*. Retrieved from http://epaa.asu.edu/ojs/issue/view/7.

Sawaki, Y. (2001). Comparability of conventional and computerized tests of reading in a second language. *Language Learning & Technology, 5*, 38-59.

Shohamy, E. (2000). *The Power of Tests: A Critical Perspective on the Uses of Language Tests.* Harlow: Pearson.

Snow, C. (2008). Crosscutting themes and future research directions. In D. August & Shanahan, T. (Eds.), *Developing Reading and Writing in Second Language Learners* (pp. 275–300). New York: Routledge.

Statistics Canada (2010). *The Daily Statistics Canada.* Retrieved July 16, 2011, from: http://www.statcan.gc.ca/daily-quotidien/100309/dq100309a-eng.htm.

Taylor, C., Kirsch, I., Eignor, D. & Jamieson, J. (1999). Examining the relationship between computer familiarity and performance on computer-based language tasks. *Language Learning, 49*, 219-274.

Toronto District School Board. (n.d.). *Facts and Figures.* Accessed March 24, 2012. http://www.tdsb.on.ca/communications/tdsbfacts.html.

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

Yu, G. (2010). Effects of presentation mode and computer familiarity on summarization of extended texts. *Language Assessment Quarterly, 7*, 119-136.