Comparing First-Term Students' English Language Proficiency at a Canadian Polytechnic Institute

Nathan J. Devos
British Columbia Institute of Technology

Abstract

Canadian post-secondary classrooms are linguistically diverse. This diversity comes from immigration, more access to higher education for marginalized and minority groups, and international student recruitment. An institute's language prerequisites serve as tools to help admissions decide who is linguistically prepared for study in English. Despite these language prerequisites, some students still need support with their language development to succeed in intense, two-year diploma programs at polytechnic institutes. The objective of this study, therefore, was to explore first-term students' language proficiency and compare differences based on language and education variables. The long-term goal of the project is to create local language diagnostics to identify and support students during their diploma programs. Participants were recruited from four different programs at a large polytechnic institute in western Canada. Reading, writing, and vocabulary tests were accompanied by a 15-item survey about language and education backgrounds. In total, 437 first-term students participated in the tests and survey. The results suggest that there are statistically significant differences between the first-term participants with an English background and those who do not have an English background. In addition, the range of scores on the tests is wider for participants with non-English backgrounds. I conclude that this range cannot be captured by language prerequisites pre-admissions. With locally developed, post-admission language assessments, diploma-crediting institutes could identify students at the lower end of the range and allocate necessary resources to provide them with early and individual language support so that they can succeed in their programs.

Background

Canadian post-secondary classrooms are increasingly linguistically diverse. This diversity reflects Canada's already multicultural society. Canada's pluralistic society alone gives reason for multicultural, multinational, and multilingual post-secondary classrooms, filled with domestic students who have rich and complex language backgrounds, facilitated by waves of immigration over the decades (Fox & Artemeva, 2017). In addition, higher education has become more accessible to minority and marginalized groups (Read, 2016). In Canada, for instance, federal and provincial governments are funding post-secondary institutes to better support Aboriginal students, students in government care, and students with accessibility challenges so they can attend and graduate with higher education credentials.[1] Finally, international student recruitment has grown almost 120% from 2010 to 2017 (ICEF Monitor, 2018). By the end of 2017, almost half a million international students were enrolled in institutions and schools across Canada. These factors have led to richly diverse language differences in post-secondary classrooms;

[1] However, whether these broad support programs can have the right effect is debatable. Fox and Artemeva (2017), for instance, have reservations about the generic nature of academic success programs and writing centres. Meanwhile, Gallop and Bastien (2016) question the success of programs for Aboriginal students in Canadian post-secondary institutes.

however, this linguistic diversity, which should be treated as a major asset to internationalization and globalization, is blanketed by one major issue—the language of instruction is strictly English.

As a result, Canadian post-secondary institutes largely require applicants to meet English language prerequisites to confirm that they have the language abilities essential to take on the demands of higher education in English. Mostly, post-secondary institutes look for academic achievements and credentials met at various levels of education or scores reached on global standardized tests. These requirements provide evidence of some background of education in English; however, education background, such as the number of years a non-native speaker spends in an English-speaking high school or studying for a standardized test, may not provide enough evidence about stages of language proficiency development.

Read (2016) states that despite second language (L2) students meeting language prerequisites, institutes still cannot assume all students entering post-secondary programs are "adequately prepared to cope with the language and literacy demands of degree studies through the medium of English" (p. 4). Evidence for this can be found, for example, in Fox (2005). Fox investigated whether the time L2 students spend in an English-medium high school makes a difference in academic achievement in the first years of post-secondary. She found that only about 15% of the students who had spent three, four, or five years in an English-medium high school were on track to meet minimum post-secondary standards. She concludes, therefore, that language residency requirements should be revisited as a prerequisite for post-secondary admission.

Furthermore, Murray (2018) points out that many students still struggle with language despite meeting entry requirements on standardized tests. He argues that gatekeeping tests such as IELTS and TOEFL "are being misused in that entry thresholds are being set too low, or that the tests are not really measuring what they need to be measuring or are only measuring part of what they need to measure" (p. 6). This may be in part because to achieve desired test scores, some students engage in short yet intense bouts of language learning of the four main skills so they can successfully achieve minimum bands commonly required for undergraduate study (e.g., IELTS 6.0) (Mueller, 2015). However, short and intense test preparation does not foster the larger, more holistic task of long-term language proficiency development and preparation for post-secondary instruction in English.

Rather, substantial and intense input and output is needed for learners to not only traverse through the early stages of acquisition (Buyl & Housen, 2015; Pienemann, 1998), but also to allow students to gain linguistic competence through "interaction and engagement with genres and registers through which content knowledge is expressed" (Llinares, Morton, & Whittaker, 2012, p. 17). This gaining of competence includes acquiring the basic vocabulary and grammar needed to engage with technical content, spontaneously conversing within the social and learning space of a post-secondary institution, expressing complex thoughts and ideas in writing, and focusing on advanced levels of academic literacy (Llinares, Morton, & Whittaker, 2012). In sum, education background prerequisites such as test scores, years in an English-speaking high school, or credit points from a previous post-secondary institute may not be good indicators of the range

of language abilities that students bring to the classrooms, nor are they good predictors of academic success for post-secondary study.

To understand students' language abilities in English, it may be beneficial to explore their language backgrounds. Research in second language acquisition (SLA) has shown that individual learner differences affect the rate and pathway of language development (Lightbown & Spada, 2013). Variables such as age of initial language acquisition, length of stay in the target language country, first language (L1), and language upbringing are all factors that have been rigorously explored in language research (Bifuh-Ambe, 2011). For instance, research shows that somewhere between the ages of 6 and 17 people slowly lose the ability to learn an L2 implicitly (DeKeyser, 2000; Hartshorne, Tenenbaum, & Pinker, 2018). The quality and quantity of language input and output are important during this "critical period" if native-like proficiency is the ultimate goal. Although the number of years spent in a target-language country is not linear to higher levels of language proficiency, the length of stay in a target language country can also be a factor in the acquisition of Basic Interpersonal Communication Skills (BICS) and Cognitive Academic Language Proficiency (CALP) (Cummins, 1981). The length of stay may prepare students in their BICS, while their exposure, language training, as well as quality and quantity of input and output can prepare students for their CALP, which is considered important for success in academic situations (Baker, 2011; Lyster, 2007). The age of initial arrival, motivation, and aptitude are also confounding variables that effect increased language proficiency, especially in adult learners (DeKeyser, Alfi-Shabtay, & Ravid, 2010). Finally, L1 background and confidence in the language of instruction are likely important factors for students' post-secondary success (Bifuh-Ambe, 2011). L1 exposure influences L2 abilities as learners "draw on the patterns of other languages they know as they try to discover the complexities of the new language they are learning" (Lightbown & Spada, 2013, p. 57), while Telbis, Helgeson, and Kingsbury (2014) note that international students who are high in their language confidence are also high in their confidence to complete their program. Clément, Baker, and MacIntyre (2003) also suggest that language confidence is connected to how willingly and frequently people use the language. Understanding language backgrounds may, therefore, provide valuable insights into how to best support students who are struggling to meet post-secondary standards.

Local, post-entry language diagnostics may be the institute's best early indicator of the range of first-term students' language abilities. Research on early alert in higher education suggests that student support services should begin within the first three weeks and up to six weeks into the semester (Hanover Research, 2014). However, instructors often first notice students who are at the most risk of falling behind or failing because of their language skills later than this period of time. Support for post-entry diagnostic assessment comes from Fox, Haggarty, and Artemeva (2016), who suggest:

> The global movement of students, the linguistic and cultural diversity of university classrooms, and mounting concerns about retention and program completion have prompted the increased use of post-entry diagnostic assessment, which identifies students at risk and provides them with academic support. (p. 43)

It should be made clear that the purpose of post-admission language assessment should be to function as a local test that provides an early alert for those students who may be in danger of struggling or failing in their programs. Fox, von Randow, and Volkov (2016) agree:

> There has been a dramatic increase in the use of post-entry diagnostic assessment to identify entering, undergraduate students at-risk and provide concomitant targeted supplementary learning support early enough in their initial university studies to prevent initial failure and foster academic success. (p. 267)

Ginther and Yan (2017) similarly note that "as entry-level language proficiency requirements fall, requirements for post-entry language proficiency testing and enrollment in EAP language support programs tend to increase" (p. 278). The diversity of Canadian society and the multiple pathways into post-secondary programs suggest that testing *all* incoming students is necessary.

Read (2016) recommends testing all incoming students because social diversity does not provide clear lines between "domestic" and "international" or "native speakers" and "non-native speakers," thus assessing students post-admissions based on these divisions is problematic and may be discriminatory. Knoch, Elder, and O'Hagan (2016) additionally argue that as access to higher education increases, students who may, based on pre-entry student records, be considered "domestic" students from either English or non-English backgrounds "may also be linguistically at risk in their academic study—even more so in some cases than their international counterparts" (p. 24). Therefore, any local post-admission language testing should include all enrolled students, without inclusion or exclusion criteria.

The purpose of the present research project, therefore, was to assess all first-term students' English language proficiency and investigate to what extent polytechnic students' education and language backgrounds affected the differences in their test scores. The goals were to explore the diversity of the language and education backgrounds of first-term students and determine the range of abilities on four different language tests and if there were significant differences between students based on these variables.

Because of the growing diversity and multiple pathways to meet language prerequisites at polytechnic institutes, research about the general English language proficiency of first-term students post-admission can provide empirical insights into language differences and range of abilities. Using a tailored instrument to assess all students regardless of their residency status, this study sought to contribute more empirical evidence to discussions around language abilities, standards, and thresholds at the institute. The long-term goal of this project was to initiate the first phase of a language diagnostic and support model so students who may struggle academically can be identified early and properly assisted during the course of their studies.

The research questions addressed in the present study are:

1. How diverse are first-term students at a polytechnic institute in their language and education backgrounds?
2. What differences exist in general English reading, writing, and vocabulary between groups of first-term students based on their language and education backgrounds?

Data collection took place in the first two weeks of September 2018, during the students' orientation week. The institute's Research Ethics Board also approved this study in October 2017.

The Study

Setting

The polytechnic institute where this study took place is on the west coast of Canada. In 2016-2017, 18,755 full-time day school students were enrolled in its programs, with 15,039 students enrolled at the campus where the research took place. Full-time students are registered in four different areas: technology, trades and technical studies, technology degrees, or apprenticeships.

Four programs from four different schools participated in the study: Financial Management (FinMan), which includes Finance, Financial Planning, and Accounting, as well as Risk and Insurance Management; Electrical and Computer Engineering Technology (ECET); Architectural and Building Technology (ABT); and Computer Systems Technology (CST). These programs were chosen because they are relatively large programs (their first-term population size was about 770 students), they represent four different areas of study, and they had a representative overall percentage of international enrolment (18%).[2]

In 2016-2017, 1,092 international students had enrolled in full-time programs at the institute. Fifty-two percent of the international students study in the School of Business, 11% in the School of Computing & Academic Studies, and 16% in the School of Construction & the Environment. In 2017, 51% of the first-term students had enrolled at the institute within one year of high school graduation.

Participants

This study recruited from all first-term students in the four abovementioned programs. That is, there were no inclusion or exclusion criteria. For the purpose of the study, I avoided tagging participants with preconceived labels of residency status, language background, prerequisites met, and so on. This was based on the assumption that student records can only provide limited information about factors that influence students' language abilities and because there are no clear lines between 'domestic' and 'international' in terms of their language learning levels or abilities.

Methods

This study took a descriptive, cross-sectional approach to investigate the proficiency of first-term students after they had been admitted into a program. Following a psychometric tradition, this

[2] The population size and percentage of international students are based on the target numbers by program from August 2017. Eighteen percent international student enrolment is 2% above the average total of international first-term enrollment and 1% above province-wide estimation of international students in post-secondary schools (Statistics Canada, 2016). However, due to estimated growth across the province and at the institute, 18% may now be below these averages.

quantitative study aimed to test the influence of different variables on language test scores (Nunan & Bailey, 2009). That is, this study sought to investigate whether language background or past education had a stronger impact on the language test scores of first-term students at a polytechnic institute.

Instruments

To survey and test the general English proficiency of the study participants, an online test and survey was created and hosted by the institute's Learning Management System (LMS). This software provided quiz functions that were combined to create a four-part general language proficiency test. The test included four areas: reading (30 cloze items), writing (two open tasks), vocabulary (100 multiple-choice items), and grammar (24 correct/incorrect items).[3] Participants were allowed 15 minutes for reading, 45 minutes for writing, 30 minutes of vocabulary, and five minutes for grammar. In addition, a 15-item survey was embedded as a SurveyMonkey® weblink at the end of the test to collect data on the participants' language and education backgrounds.

Survey. Participants completed a 15-item survey asking them about their language and education backgrounds. This survey helped answer the first research question: "How diverse are first-term students in their language and education backgrounds?"

Reading test. To test reading comprehension, tailored cloze tests were developed for each of the individual programs. Tailored tests for each program ensured that the reading passages were at participants' expected proficiency and maturity levels, as well as of interest to them (Brown, 2014; Grabe, 2009). Cloze tests for measuring reading ability have been used for both native speakers and non-native speakers and are able to measure reading abilities on both a sentential and intersentential level, overall meaning, or grammatical sensitivity, depending on which words have been deleted from the passage (Alderson, 2000; Trace, Brown, Janssen, & Kozhevnikova, 2017; Brown, 1993). Developing the tests consisted of following Brown's (2013) steps for developing tailored cloze tests and piloting five different versions with first-term students the previous semester. Item analyses of each version using classic test theory helped to identify the best functioning items. That is, items that showed the best item facility and discrimination for the certain blank in the passage were kept. The chosen blanks included between 30-40% content words and 60-70% function words in each test (Chae & Shin, 2015). Furthermore, a Tukey comparison of means showed no significant difference between the final versions of the tests in their development. The final versions for each of the four programs contained 30 items. Items were analyzed for internal reliability using both classic test theory and item response theory (IRT). Guttman's L2 showed ≥ 0.7, while an item analysis using Rasch modelling showed item reliability scores of ≥ 0.93.[4]

Writing tasks. Two constructed-response writing tasks were developed by a focus group of faculty members based on the idea of functional adequacy (Kuiken & Vedder, 2016). Kuiken

[3] Implicit grammar knowledge was also originally tested, but the test was low in inner reliability (KR-20 < .40) and will not be discussed in this paper.

[4] For validation of the reading tests, reading scores were correlated with VST scores, indicating a significant relationship ($r = 0.595$, $p < .001$).

and Vedder (2016) define functional adequacy as the successful fulfillment of the task as interpreted by the reader. They understand "functionally adequate" as being dependent on the task the writer has been given and how well the reader has received the writer's message. In their words, "the adequacy of the message transmitted by writer A is understood and interpreted by reader B with respect to the quantity, relation, manner, and quality of the message by A" (Kuiken & Vedder, 2016, p. 323). Therefore, the emphasis is not on language use per se, but rather on the function of the written message. This approach aligned with first-term expectations at the institute as students' writing tasks are largely technical, business, or workplace writing assignments, which focus on "getting the job done" and "clear and concise communication." To rate these two tasks, the focus group amended Kuiken and Vedder's (2016) analytic rating scale to include six elements that were scored between 0 (*incomplete*) to 5 (*very good*).[5] By using a multiple trait rating scale, I sought to collect more detailed information for each of the features of the tasks (Eckes, Mueller-Karabil, & Zimmerman, 2016; Fulcher & Davidson, 2007).

To reduce the risk of drawing general conclusions about the writing abilities of the respondents, two authentic writing tasks from different genres were developed (Eckes, Mueller-Karabil, & Zimmerman, 2016). The first task was a persuasive writing task that involved writing a blog post to peers about balancing a healthy lifestyle while engaging in the rigours of post-secondary study. The situation was a real-world scenario based on the institute's own blog that allows students to share their experiences with their peers. Expectations for the blog were collected from the blog's editor at the institute, making the task feel as "real world" as possible. This task required participants to write at least 250 words in 25 minutes. The second task was an informational task and asked participants to write an email to a fictitious boss, describing and commenting on two charts that presented the gender pay gap in Canada and western provinces. These charts were created based on numbers from a 2017 Statistics Canada report (Statistics Canada, 2017). Participants were asked to write at least 150 words in 20 minutes.

To score the written responses, four non-language expert raters were hired. Non-language expert raters were chosen because of the focus on the functional adequacy and not language use per se. According to Kuiken and Vedder (2014), language expert raters may rate writing samples that focus on functional adequacy too hard or too easy. In addition, to reduce negative rater effects outlined by Eckes, Mueller-Karabil, and Zimmerman (2016), such as leniency, central tendency, halo, or rater bias, raters were trained over a four-week period using piloted responses and the rating scales. During the marking process, efforts were taken to ensure that responses contained no identifying, biographical information. Scoring of the final responses involved a three-step process:

1. Two raters and the principal researcher followed a "divide and conquer" approach (J. Fox, personal communication, August 8, 2018). Groups of writing samples were categorized into: at-risk, average, and excellent. This was done holistically as a sorting and ranking procedure (Fulcher & Davidson, 2007). If any responses were not agreed upon, then these were openly discussed until 100% interrater agreement was met.
2. Two raters individually scored each response using the functional adequacy rating scale.
3. The parity model was used to operationalize the scores (Johnson, Penny, Gordon, & Fisher, 2005).

[5] The rating scale for the persuasive task can be found in Appendix A.

Spearman ρ was applied to measure how close the raters ranked the written responses. The results showed high levels of reliability between the raters ($\rho \geq 0.8$, $p < .05$).

Vocabulary size. To test vocabulary size, Nation's monolingual Vocabulary Size Test (VST) Version A was used. This test version is available on his website and can be used for testing and research purposes (https://www.victoria.ac.nz/lals/about/staff/paul-nation#vocab-tests). The VST can be used to measure both L1 and L2 speakers' written receptive knowledge (Elgort & Coxhead, 2016). The VST contains 100 multiple-choice questions that include the presentation of the word (i.e., the stem), plus the word used in a sentence so that the test-taker knows how the word can be used in context; this is followed by four possible answers, one correct and three distractors (Coxhead, Nation, & Sim, 2015). A reliability analysis of the 100 items showed Cronbach's $\alpha = 0.91$ and McDonald's $\omega = 0.91$.

Criticism of this test includes its development from corpora that may not be ideal for developing learners (Pinchbeck, 2016). In addition, the word frequencies that the test is based on come from the BNC / COCA, so some words may not be familiar to a Canadian test-taker. Also, the word frequencies do not provide a good measure of word difficulty, as high frequency words may be more difficult than low frequency words. Finally, because the VST is a test of written receptive vocabulary size, it can only give us a rough estimate about the test-taker's reading skills. It also does not tell us if or how well a test-taker could use these words in speaking or writing (Nation & Beglar, 2007). Further testing and development is undergoing to address these issues for future projects.

Analysis

The survey results were collected in SurveyMonkey® and exported as MS Excel files for analysis. The survey results were analyzed for frequencies and proportions. The reading and vocabulary parts of the test were collected and scored automatically by the LMS, whereas the writing samples were extracted from the LMS to be scored by the raters. All the test results were compiled in MS Excel tables for data validation, clean up, and filtering. Once these tables were completed, test results were entered into Minitab 18 for descriptive and inferential statistical analyses and WINSTEPS for Rasch analysis.

To compare differences of means between two factors, independent samples *t* tests were run. To compare differences of multiple means, one-way ANOVA was applied. When assumptions for parametric tests to compare means were not met, nonparametric equivalent tests were used. Either Cohen's *d* after *t* tests or omega squared (ω^2) after ANOVA were used to test the strength of association between the variables. Omega squared is considered by some as a more conservative measure than partial eta squared (η^2) when measuring effect sizes for ANOVA (Grace-Martin, 2019). For *t* tests, Plonsky and Oswald's (2014) suggestion for interpreting the between group effect size of Cohen's *d* in language testing research were used: 0.40 (small effect), 0.70 (medium effect), 1.00 (large effect).

To explain the range of test scores, the interquartile range (IQR) and standard deviations (SD) were analyzed. The IQR represents the middle 50% of the collected data and shows a larger

or smaller range in test scores relative to its size. Additionally, the SD explains to what extent the scores of a group differ on average from the central tendency (i.e., the average) (Green, 2013).

Results

Survey Results

In total, information from 437 participants was collected through a survey. The number of participants from each program is displayed in Table 1.

Table 1
Participants by Program

Program	*n*
Financial Management	199
Architectural and Building Technology	98
Electrical and Computer Engineering Technology	72
Computer Systems Technology	68

Language background. To answer the first research question, participants were asked about their language and education backgrounds.

Language confidence. The survey asked participants to choose which language they were *most confident* in at the time of the testing. Survey questions that ask students about their first language, native language, or mother tongue can lead to confusion or misunderstanding. Therefore, by asking participants to choose the language they felt most confident using, I hoped to mitigate misunderstandings of first, second, native, and mother tongue language acquisition because I had no opportunity to follow up on the responses (Dornyei, 2003).

Seventy-three percent of the participants (n = 319) responded that they were most confident using English, whereas 27% (n = 118) responded that they were most confident in a language other than English. The survey uncovered 22 different languages in which participants were most confident. The top four most frequent languages other than English included Mandarin (25%), Korean (22%), Vietnamese (13%), and Cantonese (7%).

Age of initial English acquisition and language upbringing. To get an idea of language upbringing, participants were surveyed on whether their parents or guardians had spoken English with them since infancy and at what age they started learning English. The survey results showed that 55% of the participants had spoken English to at least one of their parents or guardians since they were babies, while 40% said that they had not. The remaining 5% responded with "I don't know." In addition, 62% of the participants began learning English between birth and five years old, 32% started learning English between 6 and 17 years old, and the remaining 5% began learning English after the age of 17.

Number of years in Canada. The survey asked participants for how many years they had lived in Canada. The majority of the respondents had spent 11 or more years in Canada (65%). Eleven percent had spent between 5 and 10 years in Canada. Meanwhile, 24% of the respondents had spent 4 or fewer years in Canada.

Education background. Participants were also asked about their education background. Three areas were of particular interest because they related to different pathways students can take to gain entry into the institute: taking a standardized language test, years spent in an English-medium high school, and previous attendance at a post-secondary institute.

Standardized language tests. Most of the participants (75%) had not taken a standardized language test (e.g., TOEFL, CAEL, IELTS, PTE, etc.). Of the 17% (n = 75) who had written a standardized language test, 55% had written an IELTS test, 7% had written TOEFL (iBT), and another 7% had written CAEL. The remaining participants fell in the category "other" because they specified English 12 or Provincial 12 as being a standardized language test. This group was not considered in subsequent analyses. A follow-up question for participants who had taken a test was how long ago it had been since they had written the test. Sixty-seven percent said it was less than one year ago, 29% said they had taken the test 1-2 years ago, and 4% had taken it 3-6 years ago.

Years in English-speaking high school. When asked about how much time they had spent at an English-speaking high school, 39% of the participants (n = 173) had spent five or more years in high school, 26% spent five years in an English-speaking high school, and 19% spent 3–4 years in an English-speaking high school. On the other hand, 5% spent only 1–2 years in an English-speaking high school and 10% spent less than one year or no time at an English-speaking high school. In addition, when asked about where they received their English 12 prerequisites to gain entry to the institute, 69% said from a public high school in the province. Private high schools represented 7% of the responses, while international colleges within the province received 4% of the responses. Institutions outside the province, either public, private, or international, made up 6% of the total responses. The remaining 15% of the respondents responded with either "none" or "other."

Previous post-secondary experience. A possible language prerequisite at the institute includes transferring credits from language-related courses (e.g., Anthropology, Law, English Literature, Psychology, etc.) taken at other post-secondary schools; therefore, the survey asked participants about whether they had previous experience as post-secondary institutes. Almost half (49%) of the participants had not previously attended a post-secondary institute, whereas 44% had attended one post-secondary institute before taking the survey. Six percent of the respondents had attended two post-secondary institutes and the remaining 1% had attended three prior to starting at the institute.

Of those participants who had previously attended post-secondary (n = 257), 30% had been to the two largest universities in the province, and 55% had been to other institutes across the province. Another 9% had attended post-secondary elsewhere in Canada, and 6% had been to post-secondary outside of Canada.

Test Results

The following sections present the results of the tests organized by the variables identified in the survey. These results are divided into two main categories: language background variables and education background variables. The categories were created to answer the second research question: "What differences exist in reading, writing, and vocabulary proficiency between groups of first-term students based on their language and education backgrounds?" The results are presented in percentages and in mean scores for each of the three tests.

Age of initial English acquisition and language upbringing. In this section, two factors related to language upbringing will be presented: Having English spoken to the participants since infancy and age of initial English acquisition. The reason for looking at the latter as a variable comes from research that suggests age of initial acquisition is a robust factor in how L2s are learned. Figure 1 illustrates the results when looking at the first factor: the age of initial English acquisition.

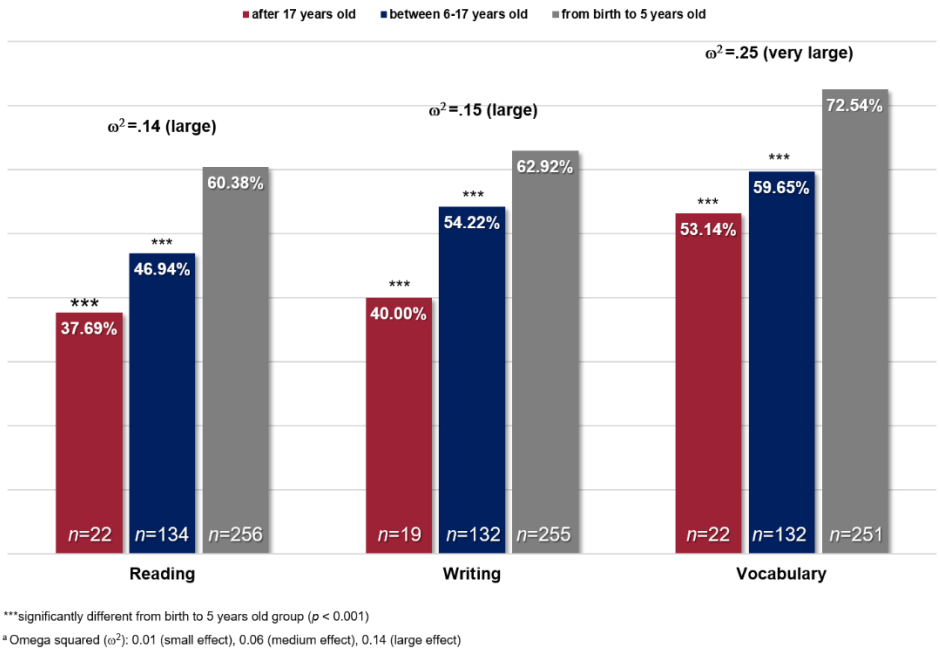

Figure 1. Comparison of Test Scores by Age of Initial English Acquisition

Figure 1 shows that there is a statistically significant difference between participants who began learning English between birth and five years old and the other two groups. Three one-way ANOVAs resulted in the following: reading $F(2, 409) = 35.69$, $p < .001$; writing $F(2, 403) = 37.90$, $p < .001$; vocabulary $F(2, 402) = 68.57$, $p < .001$. However, the Tukey pairwise comparison showed that there was no significant difference between the reading scores of the after 17-year-old group and the between 6 and 17 year-old group ($p > .05$). The difference was also not as significant between these two groups in the vocabulary scores ($p = .03$). On the other hand, the post hoc comparison showed that there were statistically significant differences between all the groups in the writing scores ($p < .001$).

Furthermore, the effect of age of initial English acquisition was the largest between the vocabulary scores ($\omega^2 = .25$), suggesting that participants' initial age of English acquisition impacts their vocabulary size more than the other two measures (Figure 1). However, the effect size is still considered large in reading and writing.

It is of interest to also look at the range of scores within these groups. To do this, I looked at the IQR and SD. These often provide a simple yet reliable way to identify which groups showed a wider range of results than others. Table 2 illustrates how the IQR is smaller in each of the tests for the birth to 5 year-old group, suggesting a tighter range around the average for the group. Although this is the largest of the groups (reading, $n = 256$), the IQR is still smaller than the other two groups. As the table shows, the range and IQR are larger for the other two groups, suggesting a wider range of scores on the tests.

Table 2
Interquartile Range and Standard Deviations by Age of Initial Acquisition

Test	Age of Acquisition	*n*	*M*	*SD*	*IQR*
Reading	from birth to 5 years old	256	60.38%	16.66	23.33
	between 6-17 years old	134	46.94%	19.21	30
	after 17 years old	22	37.69%	22	30.8
Writing	from birth to 5 years old	255	62.92%	12.38	15
	between 6-17 years old	132	54.22%	14.62	23.13
	after 17 years old	19	40%	18.91	35.83
Vocabulary	from birth to 5 years old	251	72.54%	9.4	11
	between 6-17 years old	132	59.65%	13.21	20
	after 17 years old	22	53.14%	12.67	17

In particular, the SD in the vocabulary and writing tests jump from the birth to 5 year-old group to the between 6 and 17 year-old group. This increase in the SD underscores the different range of abilities between these two groups.

In addition to age of acquisition, I analyzed the test scores between the group of students who responded "yes" to having one parent or guardian speak English to them since birth and those who responded "no" to this item. The results of this analysis are illustrated in Figure 2.

Independent samples *t* tests showed that the differences between the test scores are statistically significant: reading $t(299) = -6.79$, $p < .001$; writing $t(297) = -5.61$, $p < .001$; vocabulary $t(277) = -9.43$, $p < .001$. Using Cohen's *d* to measure the effect sizes, it is again the vocabulary scores that were most effected by whether or not the participant had been spoken to in English since infancy.

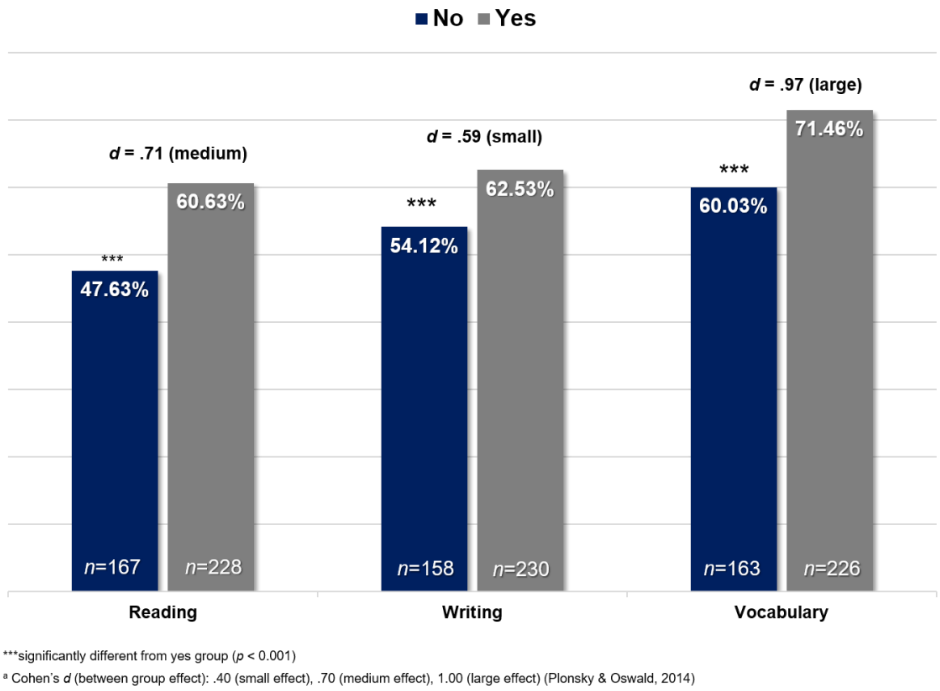

Figure 2. Comparison of Test Scores by English with Parents

The "no" group also displayed a larger range of scores in all three tests, suggested by the larger IQR and SD. For instance, the IQR of the "yes" group in their vocabulary was only 12.25, whereas it was 20 for the "no" group. The writing scores also showed much different IQR, 8.5 for the "yes" group and 14.5 for the "no" group. The difference between the groups' SD is also displayed in Figure 3.

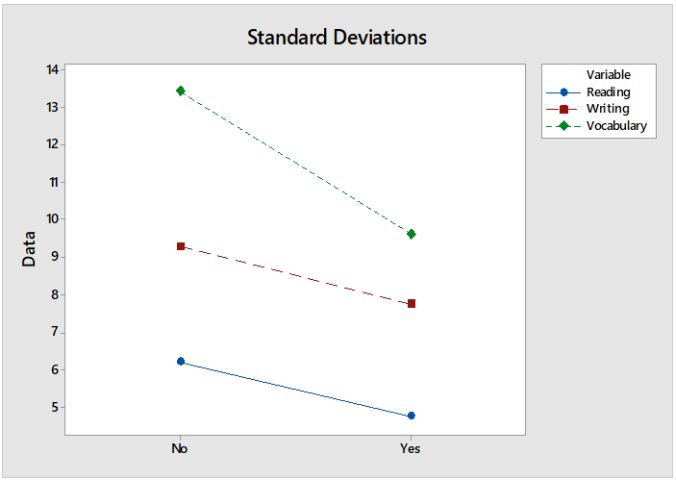

Figure 3. Line Plot of Standard Deviations of Test Scores by English with Parents

Number of years in Canada. In this study, participants were divided into three groups based on their number of years in Canada: 11+ years, 5–10 years, and ≤4 years. The test results of these three groups are presented in Figure 4.

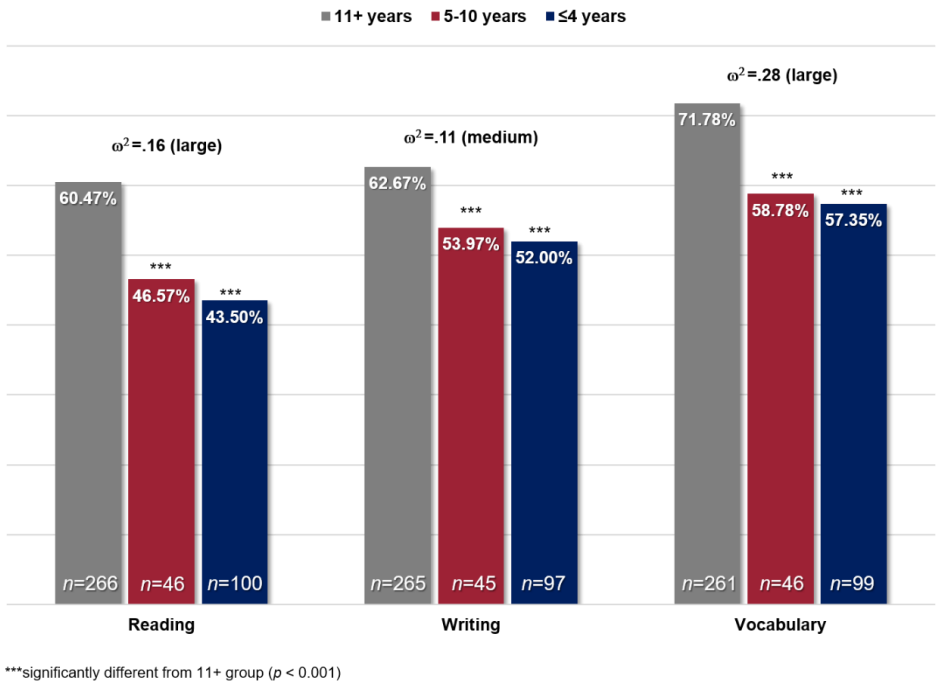

Figure 4. Comparison of Test Scores by Number of Years in Canada

The results of the one-way ANOVAs for each test showed a statistically significant difference between the 11+ years group and the other two groups in all the tests: reading $F(2, 409) = 39.09$, $p < .001$, writing $F(2, 402) = 25.50$, $p < .001$, vocabulary $F(2, 401) = 81.10$, $p < .001$. The post hoc pairwise comparison, however, indicated no statistically significant difference between the means of the test scores of the other two groups. The probability values for the post hoc *t* tests between these groups were: reading ($p = .595$), writing ($p = .706$), and vocabulary ($p = .730$).

Looking at the IQR to examine the range of the scores within these groups, we again observe wider ranges in the groups that have been in Canada for 10 or fewer years (Figures 5–7). The dispersion from the central tendency is the tightest for the vocabulary scores for the 11+ with an IQR of 10 versus the 5–10 year group with an IQR of 23.

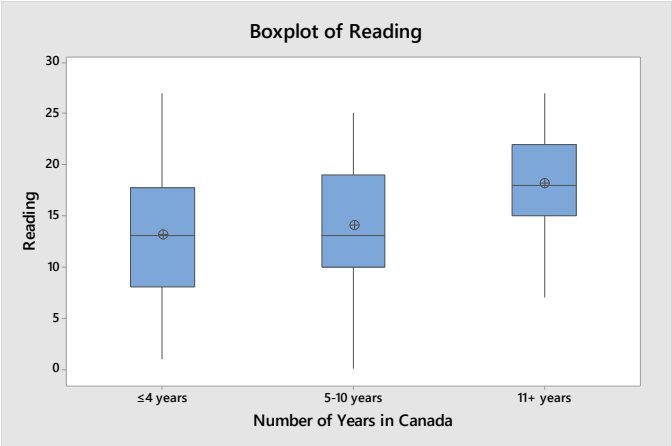

Figure 5. Boxplot of Reading Scores by Number of Years in Canada

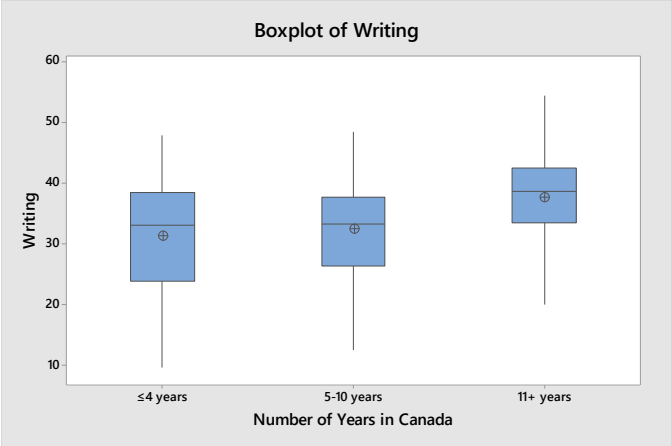

Figure 6. Boxplot of Writing Scores by Number of Years in Canada

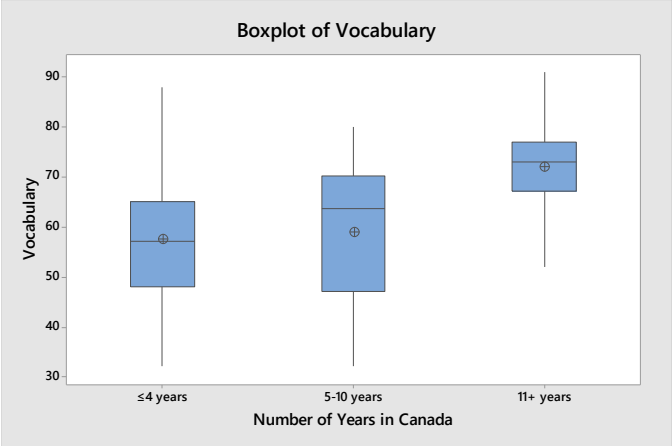

Figure 7. Boxplot of Vocabulary Scores by Number of Years in Canada

The SD also tell us about the range of test-takers' abilities, and the differences between these are clearly observed in the line plot found in Figure 8.

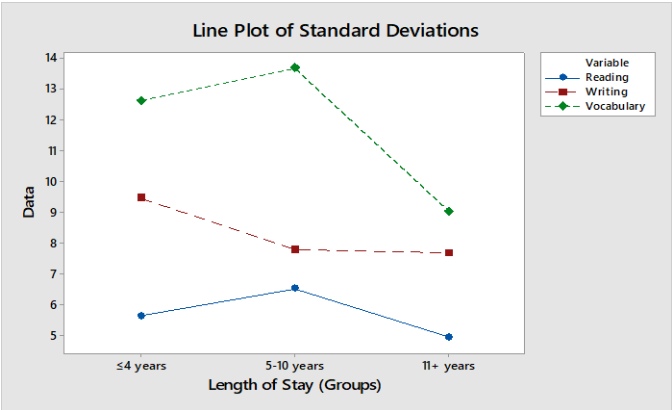

Figure 8. Line Plot of Standard Deviations by Number of Years in Canada

In sum, there is a statistically significant difference between participants who had lived in Canada for 11+ years and those who had been in Canada for 10 or fewer years. However, there was no statistical difference between the average scores if the participant had spent 5–10 years or 4 or fewer years in Canada. Furthermore, the range of scores is much larger in the 5–10 year group and the 4 or fewer group. This again suggests a wider range of abilities within these groups of participants.

Language confidence. For the purpose of this study, I considered the question of language confidence synonymous to the participant's preferred L1. I avoided asking participants for their L1, native language, or mother tongue because I expected these to be complex and may have undergone dynamic change in the participants' lives. The majority of the participants (56%) reported in the survey that they could read and write in two or more languages. This perhaps confirms the approach to ask for language confidence because sometimes bi- and trilinguals have difficulties deciding which language is their "first" language or "native" language (Baker, 2011).

Asking for language confidence also provides information about how linguistically prepared participants feel before begin their post-secondary careers. I compared the scores on the language tests between those students who felt most confident using English (the language of instruction) against those who reported feeling most confident in languages other than English (Figure 9).

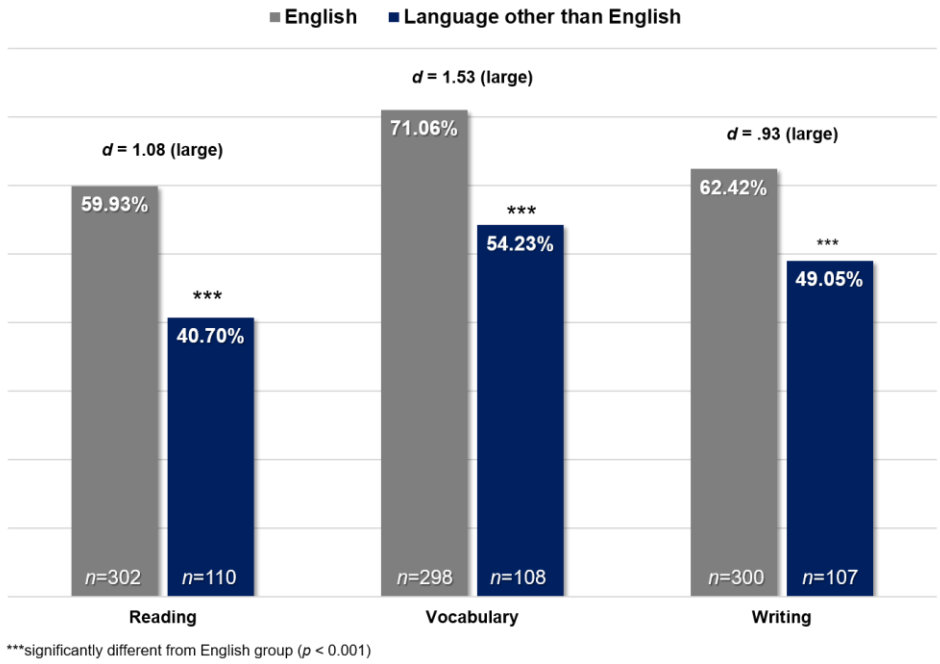

Figure 9. Comparison of Test Scores by Language Confidence

The results of independent samples *t* tests also showed the statistically significant difference between these two groups: reading $t(177) = 9.53$, $p < .001$, vocabulary $t(158) = 13.00$, $p < .001$, writing $t(160) = 7.96$, $p < .001$. As seen in the previous analyses, the largest effect size can be found between the vocabulary scores ($d = 1.53$), which is considered a very large effect size.

Similar to the other results, the boxplots found in Figures 10-12 show a clear difference in the range of scores between the two groups. The differences in the ranges were most noticeable between the vocabulary and writing scores.

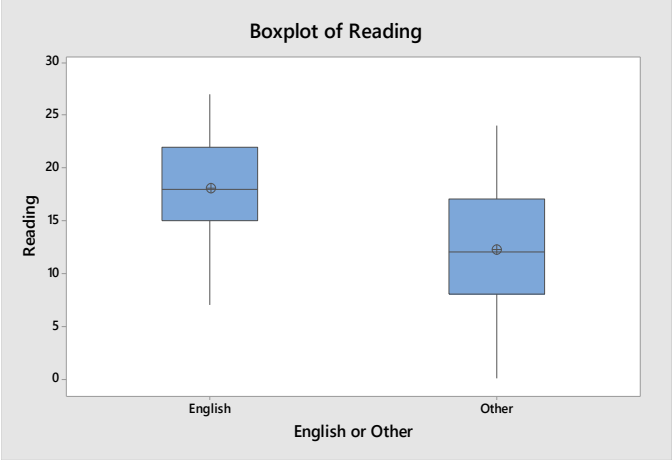

Figure 10. Boxplot of Reading Scores by Language Confidence

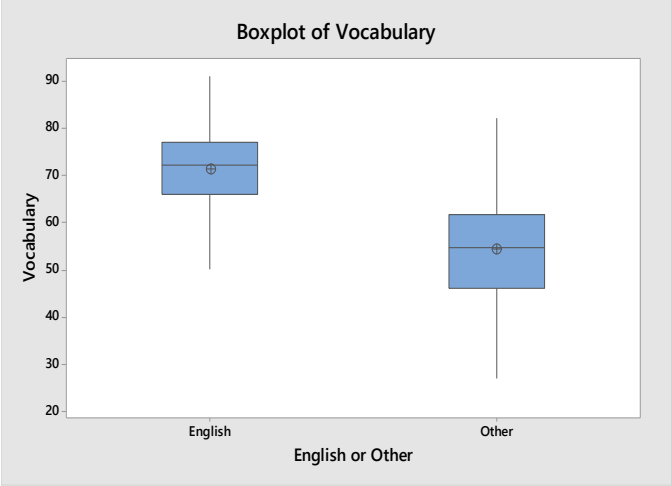

Figure 11. Boxplot of Vocabulary by Language Confidence

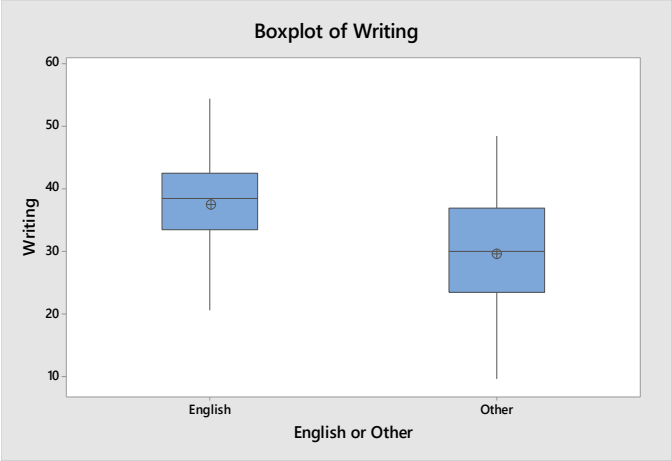

Figure 12. Boxplot of Vocabulary Scores by Language Confidence

The increased dispersion from the mean is also visible in the line plot displayed in Figure 13.

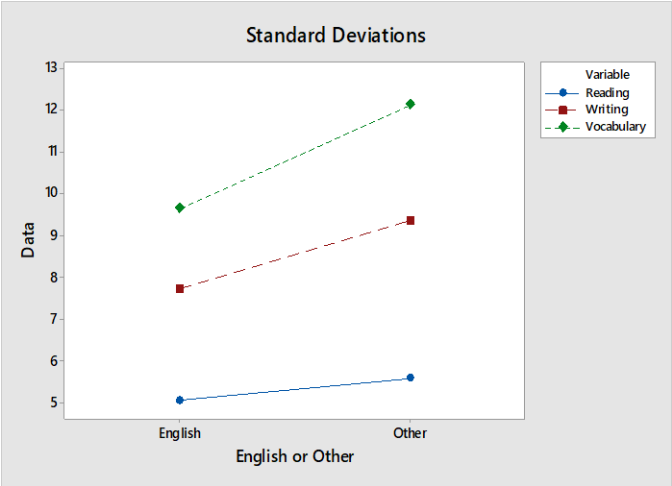

Figure 13. Line Plot of Standard Deviations by Language Confidence

The increase in SD in the vocabulary and writing scores underscore the results shown in the above boxplots and provide additional evidence for the range of abilities between the test-takers who felt confident in a language other than English.

Education Background

This project also compared the participants' education backgrounds based on their responses to the survey. The following sections present the results of comparisons between: participation in standardized language tests or not, number of years in an English-speaking high school, and previous attendance at a post-secondary institute. Noteworthy are the smaller effect sizes of these variables on the differences between the scores in comparison to the language background variables as well as the range differences between groups, which look different from the results of the previous sections.

Standardized language tests. As a possible language prerequisite to attend the polytechnic institute, participants can use their scores on a number of accredited global language tests such as IELTS, TOEFL, CAEL, and PTE. These tests are considered gatekeeping tests as the test taker must achieve a certain minimum score in order to gain entry into an institute and a program. The programs in the current study required a minimum of English 12 (67%), which was considered equivalent to: IELTS 6.5 (overall score), TOEFL 86 (overall score), CAEL 70 (overall band score), 60 (minimum on all subsets), PTE 60 (overall score). Only one of the participating programs implemented a "competitive entry" process that ranked the averages of the applicants' Math and English scores, taking the most qualified applicants based on this ranking. In the following analysis, all the participants in the "yes" group have scored equal to or above the minimums on whatever test they took.

A comparison of the test scores between those participants who had taken a standardized test and those who had not can be seen in Figure 14. The results of the *t* tests to test for significance show that the difference is the most substantial between the groups in their vocabulary scores: $t(106) = 5.36$, $p < .001$. The difference is less significant in reading, $t(106) = 3.22$, $p < .05$, and writing, $t(125) = 2.84$, $p < .05$, however. Because of the large difference in the group sizes, a nonparametric equivalent of the *t* test was run (i.e., Mann-Whitney U) and showed that the difference between the medians was significant in vocabulary ($p < .001$) but less significant in writing ($p = .005$) and reading ($p = .014$). The effect sizes of the reading ($d = .23$) and writing ($d = .34$) scores are small, while the effect on the vocabulary score ($d = .71$) is medium.

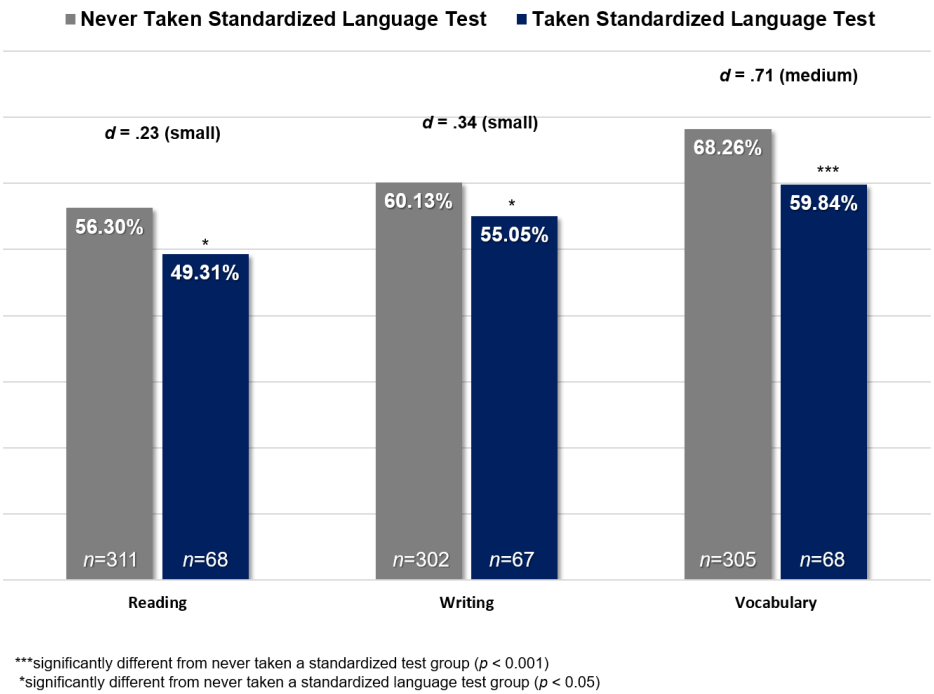

Figure 14. Comparison of Test Scores by Participation in Standardized Language Tests

An analysis of the ranges of the groups indicates that in particular the SD between the groups are closer than the SD in the language background variables (see Table 2, for example). These results are displayed in Table 3.

Table 3
Descriptive Statistics of Language Test Scores by Participation in Standardized Language Tests

Variable	Test	*n*	*M*	*SD*	*Median*	*Range*	*IQR*
Reading	No	311	56.30%	18.65	56.60%	89.9	23.3
	Yes	68	49.31%	21.1	49.93%	86.57	35.8
Writing	No	302	60.13%	14.29	63.33%	79.17	16.67
	Yes	67	55.05%	14.15	57.50%	65	21.67
Vocabulary	No	305	68.26%	12.17	70%	72	14.5
	Yes	68	59.84%	14	61%	55	20.75

The more similar SD suggest that the groups' average distances do not disperse as far from the means as when I looked at the other language background variables. This might suggest that the standardized language test variable does not show as clearly the range of abilities as the language background variables.

Number of years in English-speaking high school. The English requirements at the institute consist of completing two years of full-time education in English in an English-speaking high school and English 12. In this study, I undertook two steps while analyzing the participants' test scores and the number of years they had spent in an English-speaking high school. First, I used descriptive statistics to calculate the means of all participants based on the number of years they had spent at an English-speaking high school. The results are presented in Figure 15.

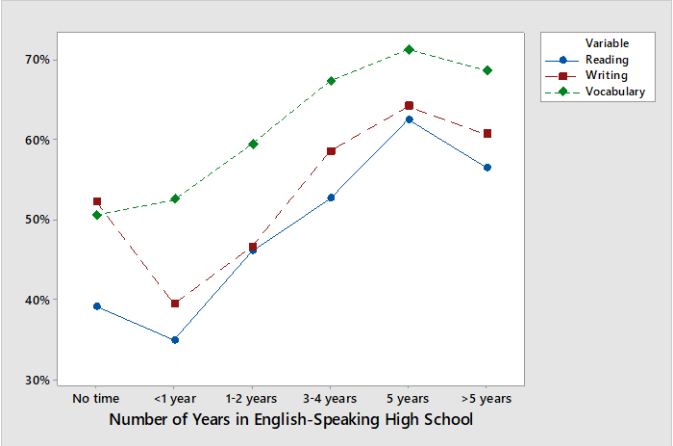

Figure 15. Line Plot of Mean Test Scores by Number of Years in an English-Speaking High School

Multiple one-way ANOVAs indicated that the differences between these scores were also significantly different ($p < .001$). The pairwise comparison showed that the differences were

most significant between the >5 year and 5-year group and the no time and <1 year group ($p \leq .001$). An analysis of the SD found the largest dispersion from the means in the <1 year, 1-2 years, and 3-4 years groups (Figure 16). The higher the SD the larger the range of scores in the group. This range in abilities was also reflected in an analysis of the IQR.

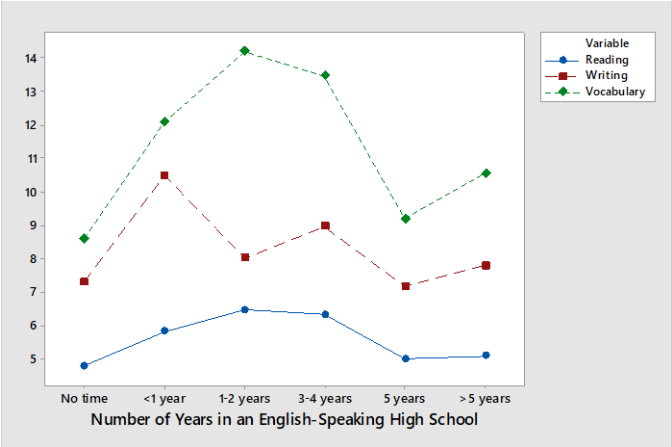

Figure 16. Line Plot of Standard Deviations by Years in an English-Speaking High School

The IQR were the highest for the 1–2 years (IQR = 33) and 3–4 years (IQR = 37) groups in reading, which compared to the 5-years group (IQR = 20). In writing, the no time (IQR = 23) and the <1 year (IQR = 25) differed the most from the 5 year group (IQR = 14). Meanwhile, in vocabulary, the <1-year group (IQR = 19) and the 3–4 year group (IQR = 18) showed the widest range of abilities, which compared to the 5 year group that had an IQR of only 9.

The second analysis of these groups involved looking at the differences in scores between non-English background participants who had chosen a language other than English as their most confident, had spent 5 or fewer years in an English-speaking high school, and lived in Canada for 10 or fewer years. To analyze the results from these remaining students, a nonparametric equivalent of the one-way ANOVA was used, the Kruskal-Wallis Test. The results presented in Table 4 show that there are no significant differences in the test scores regardless of the number of years the participant had spent in an English-speaking high school.

Table 4
Kruskal-Wallis Test of Difference Between Non-English Background Participants Based on Number of Years in an English-Speaking High School

Test	*df*	*H*	*p*
Reading (*n* = 92)	4	2.60	.626
Writing (*n* = 88)	4	5.46	.243
Vocabulary (*n* = 81)	4	5.93	.204

These findings support the abovementioned study by Fox (2005) in which she investigated the academic performance of L2 students in a Canadian university based on the amount of time they had attended English-medium high schools. She found that there was no

significant difference in academic performance between L2 students who had spent three, four, or five years at English-medium high schools. In this study, I only investigate the performance on language tests based on years in an English-speaking high school; however, the results are similar in that they suggest that the increased number of years do not improve participants' scores on the language tests.

Previous post-secondary experience. The final analysis of education background included whether or not participants had previously attended a post-secondary institute before beginning their program at the polytechnic. A possible language prerequisite for participants includes them showing the successful completion of three credits in an English, humanities, or social science related course. The applicant must receive the equivalent of 67% in that course when the program requirement is 67% in English 12. In the analysis, I looked at those students who had previously attended a post-secondary and those who had not. The null hypothesis was that there would be no statistically significant difference between the test scores. The results of this analysis are displayed in Figure 17.

The independent samples *t* tests showed that, for the reading and vocabulary tests, I had to reject the null hypothesis and accept the alternative hypothesis. That is, there were statistically significant differences between the reading and vocabulary test scores based on whether or not the participant had previously attended a post-secondary institution. The results of the independent samples *t* tests were as follows: reading $t(383) = -5.34$, $p < .001$; vocabulary $t(383)$, $p < .001$. However, I accepted the null hypothesis for the writing test scores, as the difference was not statistically significant: writing $t(380) = -1.63$, $p = .104$. Comparing to the previous variables, the effect sizes for each of these tests are much smaller, suggesting that previous post-secondary experience has a relatively small effect on the differences between the scores.

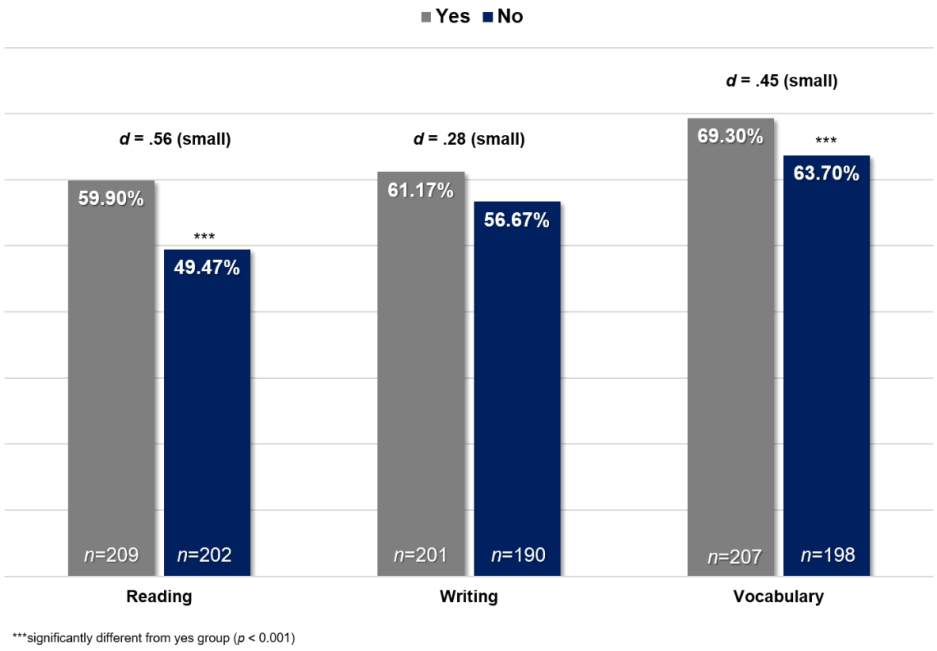

Figure 17. Comparison of Test Scores by Previous Post-secondary Experience

Regarding the range of scores in the test between the groups, the SD show minimal difference, as illustrated in Figure 18.

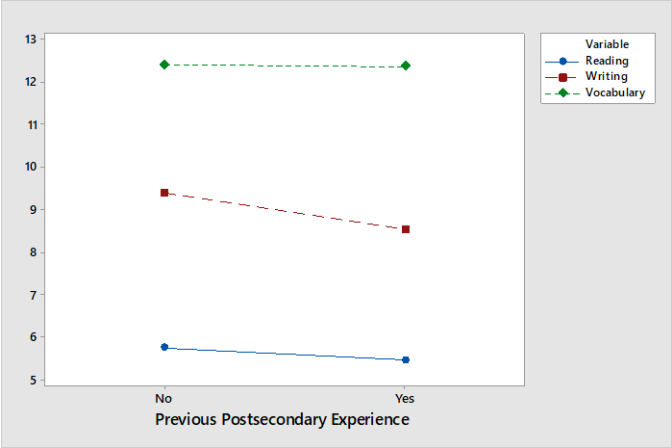

Figure 18. Line Plot of Standard Deviations by Previous Post-secondary Experience

The IQR of the groups also do not display the same differences that were observed when looking at language background variables. For example, the IQR for the reading scores was 26.63 for "no previous attendance" and 25 for "yes previous attendance." The largest difference between IQR was in the writing scores, with the "no previous attendance" (IQR = 20.52) and "yes previous attendance" (IQR = 17.92).

Discussion

The first-term participants in the current study are linguistically diverse and have complex language backgrounds, making it difficult to use dichotomous labels such as "international" or "domestic" to define who the English language learners (ELLs)[6] are and who are not. Most institutes use residency status to define "international" versus "domestic" students. However, the participants' age of initial English acquisition, whether or not their parents had spoken English with them since infancy, their length of stay, years in an English-speaking high school, and their language confidence are all complexities that are not captured by this status. Many participants in this study may have been defined at an admissions level as "domestic" students. In fact, the target average international enrolment for the participating programs was 18%, but the percentage of participants who identified a language other than English as their most confident was 27%. When looking at their language test results, some of these students could, therefore, be considered ELLs and benefit from additional language support. This underscores the notion that there are no clear lines between "international" and "English language learner," so these terms cannot be used synonymously.

The number of languages (n = 22) uncovered in this study is also noteworthy. From 118 participants, 21 languages other than English were identified. For a traditionally technical and trades institute, this language diversity is comparable to that of larger universities in Canada. For

[6] A definition of ELLs is: "[S]tudents who are learning the language of instruction at the same time as they are learning the curriculum and developing a full range of literacy skills" (Ontario Ministry of Education, 2009, p. 2).

instance, Fox (2005) reported 43 different languages other than English in first-year programs in a study of 265 participants at Carleton University, while McDonough, Neumann, and Hubert-Smith (2018) cited 30 different languages in a study of 409 participants in an EAP program at Concordia University. Having a linguistically heterogeneous student population is, on the one hand, a positive indication of internationalization and globalization that adds value to the diversification of the institute; on the other hand, it can be a daunting task for educators as each student with a different language background other than English has a different trajectory in their language progress as L1 transfer helps or hinders their development at different stages (Ellis, 2008; Jarvis, 2015). For instructors who know little about language development, stark language differences between their students may be confusing or frustrating. Educating content instructors through professional development workshops about language development and individual learner differences may help stem these frustrations from being projected back on the language learner.

The second research question in this study was about whether or not there were differences between groups of students on the tests. The findings suggest that there are statistically significant differences between students using a variety of language and education background variables. Naturally, limitations to the testing instruments exist as they only tested a small number of subskills and the test did not include important skills such as speaking and listening. Also, the sample group was self-selected because participation in the study was voluntary, meaning any particular group may be underrepresented (Dornyei, 2003). Responses about demographic information could also not be corroborated by admissions because the institute does not collect data on the language backgrounds of its students (Golder, Reeder, & Fleming, 2011). Nevertheless, the purpose of the project was to provide some empirical evidence about first-term students' language proficiency, in part, to counteract or confirm anecdotal evidence at the institute about students' abilities to use English. For instance, I did confirm that there are significant differences between students who are confident in English and those who are confident in another language, but there were no differences between the foreign languages themselves.

It may also come as a surprise to some that a post-secondary student who has successfully gained entry into a program and spent 5–10 years living in Canada might still score at the same level on a language test as a student who has spent 4 or fewer years in Canada. Because language learning is not a perfectly linear process but rather a winding journey that is psychological, cognitive, and social, the pathway to language learning may come easy for some but be seemingly never-ending for others. The correlation between length of stay in a country and language development is not necessarily a strong one (Bifuh-Ambe, 2011). Knowing that the acquisition of BICS takes two to three years, and CALPs takes five to seven years to acquire (Cummins, 1981; Roessingh, 2006), then some lower level students will likely not get close to the language levels of their native-speaking counterparts during their two-year diploma programs. Unlike longer degree programs, any language gaps become particularly magnified during intense two-year diploma programs where there is little opportunity for students to focus on the form and function of language, which is required if language development is to take place in the context of purposeful learning (Lucas, Villegas, & Freedson, 2008; Lightbown & Spada, 2013).

The results also suggest that language prerequisites cannot capture the range of language abilities students bring to their first-term programs. In the present study, this became obvious when looking at the language background variables yet became weaker when looking at the variables related to language prerequisites. This suggests that the differences that instructors experience in the classroom can be better explained by the students' language upbringing, age of initial English acquisition, language confidence, and length of stay than by whether or not they fulfilled language prerequisites or their status as "international" or "domestic" students. For example, large global proficiency tests may provide information about general academic proficiency, but as they are first-past-the-post measures, especially when only overall scores and not subskills are considered, they only inform admissions about applicants' levels of general academic proficiency. In fact, using tests that have prepared students for general academic proficiency may be even less adequate measures for technical institutes where business, technical, and other workplace writing skills are required. Knoch and Macqueen (2017) posit:

> It could therefore be argued that these tests [IELTS and TOEFL] are only suitable to make predictions about university study but not beyond to possible performance in the workplace. Any use of tests beyond the purpose they were designed for casts serious doubt on the validity of the score-based interpretations made based on the assessment. (p. 294)

This argument is underscored by the range of abilities found in the writing and reading tests, which were tailored to the expectations for workplace communication. In sum, technical and trades institutes should have a special interest in developing local tests that reflect the actual workplace language needs of the students (Knoch & Macqueen, 2017).

Looking forward, groups who had English backgrounds had smaller SD and IQR on the tests, suggesting that their scores clustered more tightly around the average. On the other hand, groups that had non-English background showed much higher SD and IQR in reading, writing, and vocabulary, suggesting that they had a much wider range of abilities. As a result, a small portion of this latter group fell far below the overall group averages. When I calculated the overall proportion of students who fell one SD or more below the group average on each of the tests, I identified about 14% (n = 70) in the vocabulary test, 19% (n = 90) in the reading tests, and 17% (n = 75) in the writing test. A follow-up study is underway to investigate whether or not these groups are actually at risk of failing, dropping out, or repeating in their programs. The range of scores highlights the need to look at students' language abilities with more local tests that are tailored to the expectations of the institute.

Implications

Students who do not have an English upbringing are expected to meet language prerequisites before entering post-secondary institutes in Canada. Many students invest resources into meeting these prerequisites, and they, and the institutes, believe they are starting their studies with the linguistic competency needed to succeed. However, there are statistically significant differences in English language proficiency between students with an English background and those who do not. As a result, some students in this study lagged far behind their English-speaking peers in writing, vocabulary, and reading skills. Therefore, if institutes could improve local, post-

admission assessment screening and diagnostic tools to identify these students both fairly and early, then essential resources could be properly allocated to them and they can receive the support they need to succeed academically. Regardless of the accuracy and specificity of locally created diagnostic tools, any language support model *must* be directly coupled with the specific needs of the learners. A precise and individual support model may be worth pursuing. For instance, similar to "precision medicine" in healthcare (Timmerman, 2013), a "precision language education" model could seek to customize language teaching to individual students in short yet intense diploma programs. This model could include options of: small-group tutorials, one-on-one advising, online modules, peer learning, and others. The selection of these methods would be based on not only the students' language learning profile, including consideration of their language and cultural background, but also on the needs of their specific programs (e.g., ESP and ESAP). Doing this might vastly improve ELL's success and learning experience in diploma programs.

Acknowledgements

This project was supported by BCIT's Institute Research Funds and the Professional Development committee in SoCAS. I would like to thank the research assistants, Jessica Wang, Katie Conconi, and Serianna Schroh, for their tireless efforts. Also, I would like to acknowledge all my great colleagues who graciously volunteered their time.

References

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

Baker, C. (2011). *Foundations of bilingual education and bilingualism* (5th ed.). Toronto, ON: Multilingualism Matters.

Bifuh-Ambe, E. (2011). Postsecondary learning: Recognizing the needs of English language learners in mainstream university classrooms. *Multicultural Education, 13*(19), 13-19. Retrieved from https://files.eric.ed.gov/fulltext/EJ955932.pdf

Brown, J. D. (1993). What are the characteristics of natural cloze tests? *Language Testing, 10*(2), 93-116. https://doi.org/10.1177/026553229301000201

Brown, J. D. (2013). My twenty-five years of cloze testing research: So what? *International Journal of Language Studies, 7*(1), 1-32. Retrieved from http://www.ijls.net/sample/71-1.pdf

Brown, J. D. (2014). *Testing in language programs: A comprehensive guide to English language assessment*. J D Brown Publishing.

Buyl, A., & Housen, A. (2015). Developmental stages in receptive grammar acquisition: A Processability Theory Account. *Second Language Research, 31*(4), 1-28. https://doi.org/10.1177/0267658315585905

Chae, E.-Y., & Shin, J.-A. (2015). A study of a timed cloze test for evaluating L2 proficiency. *English Teaching, 70*(3), 117-135. https://doi.org/10.15858/engtea.70.3.201509.117

Clement, R., Baker, S. C., & MacIntyre, P. D. (2003). Willingness to communicate in a second language: The effects of context, norms, and vitality. *Journal of Language and Social Psychology, 22*(2), 190-209. https://doi.org/10.1177/0261927x03022002003

Coxhead, A., Nation, P., & Sim, D. (2015). Measuring the vocabulary size of native speakers of English in New Zealand secondary schools. *New Zealand Journal of Educational Studies, 50*(1), 121-135. https://doi.org/10.1007/s40841-015-0002-3

Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In California State Department of Education (Ed.), *Schooling and language minority students: A theoretical framework* (pp. 3-49). Los Angeles, CA: Evaluation, Dissemination and Assessment Center, California State University, Los Angeles. Retrieved from https://files.eric.ed.gov/fulltext/ED249773.pdf

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*, 499-533. https://doi.org/10.1017/s0272263100004022

DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics, 31*(3), 413-438. https://doi.org/10.1017/S014276410000056

Dornyei, Z. (2003). *Questionnaires in second language research: Construction, administration, and processing*. New York, NY: Routledge.

Eckes, T., Mueller-Karabil, A., & Zimmerman, S. (2016). Assessing writing. In D. Tsagari, & J. Banerjee, *Handbook of Second Language Assessment* (pp. 147-164). Boston, MA: DeGruyter.

Elgort, I., & Coxhead, A. (2016). The Vocabulary Size Test in the Pacific Rim: Description, application, and use. In V. Aryadoust, *Trends in language assessment research and practice: The view from the Middle East and the Pacific Rim* (pp. 286–299). Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford, UK: Oxford University Press.

Fox, J. (2005). Rethinking second language admission requirements: Problems with language-residency criteria and the need for language assessment and support. *Language Assessment Quarterly, 2*(2), 85-115. https://doi.org/10.1207/s15434311laq0202_1

Fox, J., & Artemeva, N. (2017). From diagnosis toward academic support. *ESP Today, 5*(2), 148-171. https://doi.org/10.18485/esptoday.2017.5.2.2

Fox, J., Haggarty, J., & Artemeva, N. (2016). Mitigating risk: The impact of a diagnostic assessment procedure on the first-year experience in engineering. In J. Read, *Post-Admission Language Assessment of University Students* (pp. 43-65). New York, NY: Springer.

Fox, J., von Randow, J., & Volkov, A. (2016). Identifying students at-risk through post-entry diagnostic assessment. In *Trends in language assessment research and practice: The view from the Middle East and the Pacific Rim* (pp. 266-285). Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. London, UK: Routledge.

Gallop, C. J., & Bastien, N. (2016). Supporting success: Aboriginal students in higher education. *Canadian Journal of Higher Education, 46*(2), 206-224. Retrieved from http://journals.sfu.ca/cjhe/index.php/cjhe/article/view/184772

Ginther, A., & Yan, X. (2017). Interpreting the relationship between TOEFL iBT scores and GPA: Language proficiency, policy, and profiles. *Language Testing, 35*(2), 271-295. https://doi.org/10.1177/02655322177040

Golder, K., Reeder, K., & Fleming, S. (2011). Determination of appropriate IELTS writing and speaking band scores for admission into two programs at a Canadian post-secondary polytechnic institute. *The Canadian Journal of Applied Linguistics, 14*(1), 222-250. Retrieved from https://journals.lib.unb.ca/index.php/CJAL/article/view/19875

Grabe, W. (2009). Teaching and testing reading. In M. H. Long, & C. J. Doughty, *The Handbook of Language Teaching* (pp. 441-462). West Sussex, UK: Wiley-Blackwell.

Grace-Martin, K. (2019). A comparison of effect size statistics. *The Analysis Factor*. Retrieved from: https://www.theanalysisfactor.com/effect-size/

Green, R. (2013). *Statistical analysis for language testers*. New York, NY: Palgrave.

Hanover Research. (2014). *Early alert systems in higher education*. Washington, DC: Hanover Research. Retrieved from https://www.hanoverresearch.com/wp-content/uploads/2017/08/Early-Alert-Systems-in-Higher-Education.pdf

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition, 177*, 263-277. https://doi.org/10.1016/j.cognition.2018.04.007

ICEF Monitor. (2018, March 21). *Canada books 20% growth in 2017*. Retrieved from: http://monitor.icef.com/2018/03/canada-books-20-growth-2017/

Jarvis, S. (2015). The scope of transfer research. In L. Yu & T. Odlin (Eds.), *New Perspectives on Transfer in Second Language Learning*. Bristol, UK: Multilingual Matters.

Johnson, R. L., Penny, J., Gordon, B. S., & Fisher, S. P. (2005). Resolving score differences in the rating of writing samples: Does discussion improve the accuracy of scores? *Language Assessment Journal, 2*(2), 117-146. https://doi.org/10.1207/s15434311laq0202_2

Knoch, U., & Macqueen, S. (2017). Language assessment for the workplace. In D. Tsagari, & J. Banerjee, *Handbook of second language assessment* (pp. 291-307). Boston, MA: De Gruyter Mouton.

Knoch, U., Elder, C., & O'Hagan, S. (2016). Examining the validity of a post-entry screening tool embedded in a specific policy context. In J. Read, *Post-admission language assessment of university* (pp. 23-41). New York, NY: Springer.

Kuiken, F., & Vedder, I. (2014). Rating written performance: What do raters do and why? *Language Testing, 31*(3), 329-348. https://doi.org/10.1177/0265532214526174

Kuiken, F., & Vedder, I. (2016). Functional adequacy in L2 writing: Towards a new rating scale. *Language Testing, 34*(3), 321-336. https://doi.org/10.1177/0265532216663991

Lightbown, P., & Spada, N. (2013). *How languages are learned* (4th ed.). Oxford, UK: Oxford University Press.

Llinares, A., Morton, T., & Whittaker, R. (2012). *The roles of language in CLIL*. Cambridge, UK: Cambridge University Press.

Lucas, T., Villegas, A. M., & Freedson, M. (2008). Linguistically responsive teacher education: Preparing classroom teachers to teach English language learners. *Journal of Teacher Education, 59*, 361-373. https://doi.org/10.1177/0022487108322110

Lyster, R. (2007). *Learning and teaching languages through content: A counterbalance approach*. Amsterdam, NL: John Benjamins.

McDonough, K., Neumann, H., & Hubert-Smith, N. (2018). How accurately do English for academic purposes students use academic word lists? *BC TEAL Journal, 3*(1), 77-89. Retrieved from https://ojs-o.library.ubc.ca/index.php/BCTJ/article/view/293

Mueller, A. (2015). The difference in error rate and type between IELTS writing bands and their impact on academic workload. *Higher Education Research & Development, 34*(6), 1207-1219. https://doi.org/10.1080/07294360.2015.1024627

Murray, N. (2018). University gatekeeping tests: What are they really testing and what are the implications for EAP provision? *JACET Journal, 62*. Retrieved from http://wrap.warwick.ac.uk/104812/

Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13. Retrieved from http://jalt-publications.org/tlt/issues/2007-07_31.7

Nunan, D., & Bailey, K. M. (2009). *Exploring second language classroom research: A comprehensive guide*. Boston, MA: Heinle.

Ontario Ministry of Education. (2009). *Supporting English Language Learners*. Retrieved from http://www.edu.gov.on.ca/eng/document/esleldprograms/guide.pdf

Pienemann, M. (1998). *Language processing and second language development*. Amsterdam, NL: John Benjamins.

Pinchbeck, G. G. (2016). Towards appropriate L1 & L2 scales of English vocabulary in English dominant contexts: Vocabulary test item difficulty indicates lexical sophistication scale and definition of "word". In R. Waring, L. Anthony, C. Browne, & T. Ishii, *Vocabulary Learning and Instruction* (Vocab@Tokyo: Current Trends in Vocabulary Studies (pp. 72-73). Tokyo: VLI: A Journal of Vocabulary and Research. Retrieved from http://vli-journal.org/vocabattokyo/vocabattokyo_handbook_2016.pdf

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878-912. https://doi.org/10.1111/lang.12079

Read, J. (2016). Some key issues in post-admission language assessment. In J. Read, *Post-Admission Language Assessment of University Students* (pp. 3-20). New York, NY: Springer

Roessingh, H. (2006). BICS-CALP: An Introduction for Some, a Review for Others. *TESL Canada Journal, 23*(2), 91-96. https://doi.org/10.18806/tesl.v23i2.57

Shapiro, S., Farrelly, R., & Tomas, Z. (2014). *Fostering international student success in higher education*. Alexandria, VA: TESOL International Association.

Statistics Canada. (2016, October 20). *International Students in Canadian Universities*. Retrieved from: https://www150.statcan.gc.ca/n1/pub/81-599-x/81-599-x2016011-eng.htm

Telbis, N. M., Helgeson, L., & Kingsbury, C. (2014). International students' confidence and academic success. *Journal of International Students, 4*(4), 330-341. Retrieved from https://files.eric.ed.gov/fulltext/EJ1054787.pdf

Timmerman, L. (2013, February 13). *What's in a name? A lot, when it comes to 'precision medicine'*. Retrieved from: https://xconomy.com/national/2013/02/04/whats-in-a-name-a-lot-when-it-comes-to-precision-medicine/

Trace, J., Brown, J. D., Janssen, G., & Kozhevnikova, L. (2017). Determining cloze item difficulty from item and passage characteristics across different learner backgrounds. *Language Testing, 34*(2), 151-174. https://doi.org/10.1177/0265532215623581

Appendix A: Rating Scale: Persuasive Task (Adapted from Kuiken & Vedder, 2016)

Content: Is the number of information units provided in the text adequate and are they relevant?					
0	**1**	**2**	**3**	**4**	**5**
Zero **relevant information units** are included or they're included but have no connection to healthy living.	Only **1-2 somewhat relevant information units** are included and the examples are somewhat relevant.	**1 information unit** and the examples **have to do with the healthy living**.	**2 different information units** are included and the examples **have to do with the healthy living**.	**3 different information units** are included in the response and all the examples **have to do with the healthy living.**	**More than 3 different information units** are included in the response and **all the examples have to do with healthy living.**

Persuasion: To what extent has the message been persuasive?					
0	**1**	**2**	**3**	**4**	**5**
The message is **not persuasive**. The reader is **not convinced** by any of the writer's suggestions for a healthy lifestyle.	The message is **barely persuasive**. The reader is **only just convinced that any of the** writer's suggestions for a healthy lifestyle would be helpful.	The message is **somewhat persuasive**. The reader is **somewhat convinced** by some of the writer's suggestions for a healthy lifestyle.	The message is **mostly persuasive**. The reader is **mostly convinced** by some of the writer's suggestions for a healthy lifestyle.	The message is **persuasive**. The reader is **convinced** by the writer's suggestions for a healthy lifestyle.	The message is **very persuasive**. The reader is **completely convinced** by all of the writer's suggestions for a healthy lifestyle.

Genre: How many elements of a blog post has the author used to make it successful?					
0	**1**	**2**	**3**	**4**	**5**
The response reads like **anything but** a blog post (e.g., academic writing, high school essay, text message, etc.)	The response **doesn't really read** like a blog post to peers because it **doesn't include any** of the following elements: (1) speaks from the writer's point of view, (2) shares personal life events, (3) talks about your day-to-day activities, (4) invites comments from others.	The response **reads somewhat** like a blog post to peers because it includes **similar elements** to: (1) speaks from the writer's point of view, (2) shares personal life events, (3) talks about your day-to-day activities, (4) invites comments from others.	The response **reads like a blog post to peers** because it includes **one or two elements** of: (1) speaks from the writer's point of view, (2) shares personal life events, (3) talks about your day-to-day activities, (4) invites comments from others.	The response reads like a **blog post to peers** because it includes **three elements** of: (1) speaks from the writer's point of view, (2) shares personal life events, (3) talks about your day-to-day activities, (4) invites comments from others.	The response really reads like a **blog post to peers** because it includes **all four elements**: (1) speaks from the writer's point of view, (2) shares personal life events, (3) talks about your day-to-day activities, (4) invites comments from others.

Tone: How well does the message match the feeling that's required for a blog post?					
0	**1**	**2**	**3**	**4**	**5**
The blog post is **far from feeling honest**, **positive**, and/or **not appropriate for 0000's website**.	The blog post **somehow doesn't feel honest**, **positive**, and/or **not suitable for 0000's website.**	The blog post feels **somewhat honest**, **positive**, and **not suitable for 0000's website**.	The blog post feels **mostly honest**, **positive**, and **suitable for 0000's website**.	The blog post feels **honest**, **positive**, and **professional for 0000's website**.	The blog post feels **very honest**, **positive**, and **professional for 0000's website**.

Comprehensibility: How much effort is required to understand the text's purpose and ideas?					
0	**1**	**2**	**3**	**4**	**5**
The response is **not at all comprehensible.** Ideas are unclearly stated and you can't understand large parts despite your efforts.	The response is **barely comprehensible.** The message is **not clear** and you **struggle to understand** the writer's ideas.	The response is **somewhat comprehensible.** **Multiple parts are hard to understand** at a first reading because of sentence structure, grammar, and/or spelling errors.	The response is **comprehensible**. **Some parts might need to be read twice** because of sentence structure, grammar, and/or spelling errors	The response is **easily comprehensible** and reads fluently. Some parts may be unclear but are **understood without too much effort**.	The response is **very easily comprehensible** and **interesting to read**. You only need to **read the message once** and **do not need to reread any sentences** for understanding.

Cohesion: How well has the author linked ideas together using a mix of phrases and words?					
0	**1**	**2**	**3**	**4**	**5**
The message is not cohesive and **ideas are not linked together at all.** No transitional phrases, words, or conjunctions are used, and ideas seem completely unrelated.	The message is barely cohesive and **only uses simple transition words to link ideas together (e.g., "and" , "but")** Ideas are not well linked by transitional words, phrases, or conjunctions; these are rarely used.	The message is somewhat cohesive and **links a couple of sentences in the blog post.** Some transitional words are used, but they are mostly conjunctions.	The message is **cohesive and links a few sentences in the blog post.** It is limited in its use of transitional phrases, parallelism, transitional words, conjunctions, and sequence words to connect sentences.	The message is cohesive and **links sentences together in parts of the blog post.** It uses transitional phrases, parallelism, transitional words, conjunctions, and sequence words to connect sentences.	The message is very cohesive by linking **sentences and paragraphs** together **throughout the whole blog post.** It uses any number of transitional phrases, parallelism, transitional words, conjunctions, and sequence words to connect sentences.