

Developing and Validating a Post-Admission Screening-Diagnostic Assessment Procedure to Offer Language Support in College Diploma Programs

Nathan Devos

British Columbia Institute of Technology

Deo Nizonkiza

University of British Columbia

Sarah Lynch

University of British Columbia

Abstract

As post-secondary institutions assume more responsibility for the language abilities of their graduates, more attention is being paid to post-admission language support to enhance student success. Previous research has indicated that a post-admission language diagnostic assessment procedure, when coupled with language support services, can be an effective model in helping students meet language expectations in post-secondary settings. This paper outlines the development and validation of a screening-diagnostic assessment procedure to recommend students to language support services in college diploma programs. Our key findings suggest that students who receive a recommendation through the procedure and subsequently attend language support (LS) classes have higher communication grades than those who do not attend, and that testing vocabulary can be an effective measure for screening language abilities. These results offer validity evidence for using this procedure while ongoing research is being conducted to validate its testing measures.

Résumé

Les établissements postsecondaires se chargent de plus en plus des compétences langagières de leurs étudiants. Le soutien linguistique visant à améliorer la réussite des étudiants s'est ainsi vu accordé une importance croissante. D'ailleurs, les recherches antérieures indiquent en effet qu'une évaluation diagnostique suivie de soutien linguistique, peut être un modèle efficace pour aider les étudiants à répondre aux attentes linguistiques de leurs établissements. Notre étude porte sur le développement et la validation d'une évaluation diagnostique suivie de recommandations aux services de soutien linguistique. Nos principales conclusions suggèrent que l'évaluation du vocabulaire peut être une mesure efficace pour le dépistage des compétences linguistiques. Par ailleurs, les étudiants qui reçoivent une recommandation aux services de soutien au terme de cette évaluation et qui assistent par la suite à des cours de soutien linguistique ont une meilleure performance dans leurs cours de communication, dépassant leurs paires qui n'y assistent pas. Ces résultats, quand bien même préliminaire, nous donnent une preuve de validité de notre évaluation diagnostique. Notre recherche continue afin de confirmer la validité de notre procédure.

Developing and Validating a Post-Admission Screening-Diagnostic Assessment Procedure to Offer Language Support in College Diploma Programs

Introduction

Internationalization and influxes of immigration have prompted the need for post-admission language support in many post-secondary institutions where English is the medium of instruction (Fox, Haggerty, & Artemeva, 2016; Read, 2008, 2016). In Canada, ample evidence exists to support trends of internationalization and first-generation immigration in higher education. Indeed, Canada has become a favourite destination for international students, with the number of international students reaching a peak of over 600,000 in 2019 (Crossman et al., 2022). In addition, the number of newcomers and landed immigrants who enter college from various pathways has added to the linguistic diversity in post-secondary classrooms (Fox, 2005; Fox, Haggerty, & Artemeva, 2016; Fox, von Randow, & Volkov, 2016). At the institute where this study takes place, roughly 25% of all students consider a language other than English (i.e., the language of instruction) to be the one they are most confident using, and almost 40% use a language other than English at home (Devos, 2022). The strength of students' language abilities in English medium post-secondary instruction has recently raised questions about the extent to which students with other language backgrounds are prepared for academic study in English (Arkoudis et al., 2012; Read, 2016). In this context, scholars have observed that the increase in students with a first language other than English has created a considerable need for language support (Devos, 2019; Fox, 2005, 2015; Fox, Haggerty, & Artemeva, 2016). Consequently, diagnostic language assessments have become an indispensable tool for evaluating and recommending students to academic support services, with its central goal of identifying language strengths and weaknesses. As the focus is put on the latter, these become the baseline for pedagogical intervention (Alderson et al., 2015; see also Alderson & Huhta, 2005). For this reason, diagnostic assessment has also been referred to as learning-oriented (Alderson, 2005; Alderson et al., 2015; Fulcher, 1997; Read, 2008, 2015b).

The DELNA (Diagnostic English Language Needs Assessment) is perhaps the most widely known post-admission diagnostic language assessment in the field. Used at Auckland University, it falls within what is known as Post-Entry Language Assessment (PELA) in Australia and New Zealand, where many such tests are used. Other major PELAs include the Diagnostic English Language Assessment (DELA) at the University of Melbourne, the Measuring the Academic Skills of University Students (MASUS) used by the University of Sydney (Bonanno & Jones, 2007), and the Online Post Enrolment Language Assessment (OPELA) used at the University of Technology Sydney (Edwards et al., 2021). In 2011, 27 universities in Australia had PELAs for various student cohorts (Arkoudis et al., 2012). However, the DELA and MASUS seem to be the best-documented PELAs in Australia (Elder & Read, 2015).

The DELA was developed in the 1990s by the Language Testing Research Centre (LTRC) and became compulsory at the University of Melbourne in 2009 (Elder & Read, 2015; Ransom, 2009). It consists of reading, writing, and listening, skills perceived as needed by tertiary-level students. With the test results, students are placed in three categories, i.e., support required, support recommended, and language sufficient, and faculties are given relevant options from which to choose for their students (Ransom, 2009). Even with limitations, such as a lack of the same understanding of the test's policy

across faculties, DELA fulfills the promises of a PELA in an Australian context (Elder & Read, 2015; Ransom, 2009), with many institutions stressing the need for diagnostic assessments and clearly giving them considerable attention since 2007 (Dunworth, 2009; Read, 2015a). The DELA has an online alternate, the Academic English Screening Test (AEST)—also known as Post-Entry Assessment of Academic English (PAAL)—which is used at the University of Melbourne and other institutions in Australia. The test consists of two sections, text completion and speed-reading tasks, completed in 25 minutes but offering the same information as DELA.¹ It was developed in 2009 by the LTRC, and according to the test designers, in case more diagnostic information is needed, a writing task drawn from the DELA is offered. Being offered online and less time-consuming, this test has an advantage over DELA (Elder & Read, 2015).

MASUS is a procedure developed by the Language Centre staff of the University of Sydney in the 1990s in response to growing concerns about students' unsatisfactory literacy skills (Bonanno & Jones 2007; Paton, 2007; Scouller et al, 2008). MASUS consists of a discipline-specific writing task implemented in a two-step process wherein students are presented with input (for a few weeks) before undertaking the writing task (Bonanno & Jones 2007). MASUS has been implemented in institutions with degree programs in Pharmacy, Accounting, Architecture, Electrical Engineering, and Law (Elder & Read, 2015, p. 43). The MASUS approach has been referred to as an *embedded diagnostic assessment* (Palmer et al., 2018). MASUS is not aimed at testing subject knowledge but the ability to analyze, evaluate, and organize information on a given topic, which is the rationale behind providing background information first (Bonanno & Jones, 2007). According to the test designers, MASUS identifies students' writing strengths and weaknesses, with the aim to offer support to students who may be at risk of failure. Students' written production is measured against four criteria: (i) retrieving and processing information from provided materials, (ii) text structure and development, (iii) academic style mastery, and (iv) grammatical accuracy. Each of these four attributes is assessed on a scale, from 4 to 1, with students placed at 2 or 1 identified as "at risk" (Bonanno & Jones, 2007). Validation studies from the University of Sydney (Dyson, 2009; Holder et al., 1999; Paton, 2007) and other institutions in Australia such as the University of New South Wales (Skinner & Mort, 2009) and outside of Australia (Erling & Richardson, 2010) suggest that the procedure is reliable and valid. The procedure identifies students' strengths and limitations, and there is evidence that supports offered based on MASUS diagnostic assessment leads to improvement (Palmer et al., 2014). In addition, students overwhelmingly support the procedure, especially its embedded diagnostic nature (Palmer et al., 2018). Erling and Richardson (2010) caution, however, that it may measure one construct but not all the aspects the procedure suggests it is.

However, DELNA is likely one of the most well-known PELAs and aligns with the objectives of a diagnostic assessment (Alderson et al., 2015; Read, 2015b), with its overarching aim to evaluate prospective university students' language skills and direct them to appropriate support in case they are identified as at-risk of failing in their programs. According to the test designers, DELNA is a free low-stakes test not associated with admission (Read, 2008). DELNA's development and piloting (Elder & Erlam, 2001), validation (Elder & von Randow, 2008; Erlam & Botelho de Magalhães, 2021), and evaluations (Read, 2008, 2016) indicate that the assessment has been successfully implemented and is suitable for its purpose. It consists of two parts: (1) screening, which assesses vocabulary and speed reading and (2) diagnosis, which consists of listening to a

mini-lecture, reading academic-type texts, and writing an interpretation of a graph (Read, 2008). At the University of Auckland, DELNA is available to all students regardless of their immigration status and is mandatory in many undergraduate programs and for all PhD students (Erlam & Botelho de Magalhães, 2021; Hirsch, 2020). A key feature of this assessment, and central to any diagnostic assessment, is that it places responsibility on students to use the available supports to develop and improve their language skills and meet the academic standards required of them.

DELTA (Diagnostic English Language Tracking Assessment) is another diagnostic test used at the Hong Kong Polytechnic University and two other partner institutions, Lingnan University of Hong Kong and Hong Kong Baptist University. While DELTA shares similar features with DELNA, it was initially designed as a proficiency measure; the diagnostic component was added later. The test was developed in response to a lack of opportunities for university students to improve and track their English proficiency skills as existing tests were mainly summative in nature, prompting the need for a diagnostic assessment (Urmston et al., 2012). Piloting studies and feedback from both students and teachers led to the final version of the test. DELTA measures vocabulary, listening, reading, and grammar in a multiple-choice format. Upon taking the test, students receive a diagnostic feedback report, providing them with advice on what could be improved and indicating which actions to take. Evaluation studies suggest that DELTA indeed fulfills both diagnostic and proficiency tracking functions (Urmston et al., 2016). A study conducted by Urmston et al. (2016) shows significant improvements over one year at the university, clearly demonstrating that students who make good use of available supports develop their English proficiency skills over time. The authors acknowledge, however, that differences in computer literacy, time and motivation constraints, and the voluntary nature of using supports and seeking advice from instructors are the main limitations of DELTA.

It is in this worldwide context of supporting students post-admissions that multiple Canadian post-secondary institutes have begun offering post-entry support for English as Additional Language (EAL) students. Many of these are run by student services, such as peer mentoring, writing centres, learning commons, and conversation groups. In terms of PELAs, the Canadian Academic English Language (CAEL) Test evolved from originally being a placement test to a nationally used proficiency test (Elder & Read, 2015). At Carleton University, CAEL has been used in a hybrid model PELA with DELNA to diagnose the discipline-specific language skills of first-year engineering students (Fox, Haggerty, & Artemeva, 2016; Fox, von Randow, & Volkov, 2016). Beynen (2020) mentions that this diagnostic assessment includes a writing task based on an in-class lecture, an academic vocabulary task, a cloze-elide reading task, as well as math tasks.

At the institute where this study is set, students in multiple college-level technical and business diploma programs complete a process like the DELNA called ESTP-O (English Screening Test for Polytechnics–Online). Students participate in the ESTP-O after they have a seat in their program and shortly before the semester begins. The purpose of ESTP-O is to help identify first-term students who may be at risk in terms of their English language skills and offer them weekly, integrated language support (LS) classes to increase their chances of success. To avoid discriminating against certain groups of participants (e.g., international students, exchange students, and newcomers to Canada), the goal of the assessment is to measure the language abilities of all students in participating programs. The origins of ESTP-O began in 2018 as a research project to measure first-term students' general proficiency in grammar, vocabulary, reading, and writing in technical and business

diploma programs at the community college level (Devos, 2019). In 2019, however, portions of the test were retrofitted for a screening-diagnostic test. Participating programs had been offering integrated, non-credit-bearing LS classes in communication courses since 2009.

Before the ESTP-O, programs had various paper and computer-based methods for diagnosing language skills; therefore, some programs sought a standardized, online process for recommending students to these integrated support classes. The development of ESTP-O as an online, remote screening-diagnostic test became salient in 2020 and 2021 during COVID-19 when all programs went online and wanted a remote method to diagnose students' language before the term started. Consequently, the screening test was administered online and remotely via the institution's Learning Management System (LMS), Desire2Learn (D2L). Through this process, students recommended to LS classes gain access to an extra one to two hours of class time per week. The curriculum framework for these non-credit-bearing LS classes states that they include either in-person or online one-on-one coaching, tutorials, or small-group lessons to provide academic learning strategies to succeed in coursework. Students practise writing and speaking tasks relevant to their program as LS instructors have the flexibility to choose outcomes and activities according to their immediate needs. Lesson topics can vary by the program to accommodate students' different communication tasks and language learning is contextualized to the field of study.

A local test for screening and diagnostic purposes was chosen over acquiring existing diagnostic assessments such as DELNA, DELA, DELTA, or MASUS for three reasons. First, by considering various factors of the local context, we could develop a test fit for purpose (Dimova et al., 2020; Norris, 2012). The local context of the post-admission and support program is two-year, full-time diplomas at the community college level, so a test for this short-cycle tertiary education (SCTE) setting was required. Second, lower English entry requirements are often needed for community colleges. As an example, for an undergraduate engineering program at a research-intensive institution such as the University of British Columbia (UBC), students are required to have at least four years of education in an English-medium school and a minimum 70% in English Studies 12 (or equivalent course)² (The University of British Columbia, n.d.). However, at the institute of the current study, engineering students are required to have only two years of education in an English-medium school and a minimum of 67% in English Studies 12 (or equivalent course). Because of lower entrance requirements, some EAL and first language (L1) students face language challenges in their programs and require different assessments and remedial support (Heeren et al., 2021). Third, the target language use (TLU) domain in vocational-oriented programs focuses on business and technical communication skills. Therefore, the ESTP-O is oriented more toward testing English for professional purposes (Douglas, 2000; Knoch & Macqueen, 2020). Therefore, unlike DELNA for example, which was designed as a "general measure of academic English" (Elder & Erlam, 2001, p. 6), the ESTP-O aligns more with the "workplace community repertoire" (Knoch & Macqueen, 2020, p. 61). Furthermore, the MASUS procedure focuses on diagnosing academic literacy in writing and involves collaboration with discipline-specific instructors and raters to develop and score written responses. With over 23 programs and 1,400 students involved in ESTP-O, the development and scoring used in the MASUS procedure were not practical because the required resources outweighed the available resources to implement such a procedure (Bachman & Palmer, 1996). Therefore, the adoption of other PELAs would

likely not be considered useful or practical by stakeholders in technical and business college diploma programs.

Theoretical Framework

Language Diagnostic Assessment

The purpose of diagnostic language assessment is to identify strengths and weaknesses in specific language areas so that test takers can use the information provided by the assessment to further their own language development (Alderson et al., 2015; Read, 2015b). Unlike large-scale language proficiency testing, whereby score reporting is largely numerical and directed towards test users (e.g., college admissions, program coordinators, instructors), diagnostic language assessment is student centred and should translate into pedagogical intervention to promote language growth. Therefore, the assessment should always be coupled with some form of pedagogical support (Fox & Artemeva, 2017). Jang (2009) discusses how “skill profiles” can offer specific feedback on individual test takers’ competencies in tested skills. However, defining competency or mastery of language skills suggests a known theory and trajectory of language development, which second language acquisition research has “failed to deliver” for test-based diagnosis (Alderson, 2007, p. 21). Therefore, our diagnostic assessment development focuses on identifying key areas of language ability that may offer some insights into academic achievement in the setting of community college based on both theory and practice. The following paragraphs outline the theoretical underpinnings of the screening-diagnostic assessment, the ESTP-O. The stated purpose of the ESTP-O is to identify first-term students who may be considered at-risk in terms of their English language abilities and offer them LS services to mitigate their risk of failure. The students included in this diagnostic assessment procedure are first-term students in 23 participating SCTE programs, from Mechanical Engineering to Food Technology.

From a theoretical perspective, we base our diagnostic assessment on a theory of communicative language ability (CLA). We used Bachman and Palmer’s (1996, 2010) language model, which is reputed as the most comprehensive to date and assessment-oriented (Celce-Murcia, 2007). CLA underlies well-known frameworks for language ability, such as the Common European Framework for References (CEFR) (Council of Europe, 2020) and the Canadian Language Benchmarks (CLB) (CLB, 2012). Alderson (2005) and Alderson et al. (2015) emphasize that while a diagnostic assessment is not a proficiency test, it should be informed by a language ability framework; a view also supported by Knoch (2011) who endorses the CLA theory. The diagnostic assessment thus identifies what has developed and to what extent and what is yet to be developed and builds on the latter to further support the learners, “which is, after all, the purpose of diagnosis” (Alderson, 2005, p. 29). Our assessment consists of two components, screening and diagnosis. For the screening portion of our test, we focus on language knowledge, in particular grammatical knowledge of vocabulary, morphosyntax, and cohesion.

Screening: Grammar and Vocabulary

Our basis for testing grammar ability and vocabulary knowledge as measures of language ability was largely based on theory. Grammatical knowledge is salient in the

production and comprehension of “formally accurate utterances or sentences” (Bachman & Palmer, 1996, p. 68). Measuring grammatical knowledge can serve as a quick and efficient measure of students’ academic language ability. First, understanding morphosyntax in English requires knowledge of morphological and syntactical forms and meanings (Purpura, 2004). According to Purpura (2004), the ability to use grammar accurately comes through practice and experience. Second, knowledge of vocabulary is considered a useful indicator of test takers’ language proficiency levels and has been discussed in multiple studies (Laufer & Levitzky, 2018; Nation, 2006, 2013; Read & Chapelle, 2001; Zareva et al., 2005). The DELNA also uses vocabulary in the screening portion of its test (Elder & Erlam, 2001). Laufer and Levitzky (2018) point out that success in L2 reading, writing, and general language proficiency can depend on a learner’s vocabulary size. Overall, vocabulary knowledge is considered a strong predictor of language ability and improves concurrently with language growth (Nation, 2013; Zareva et al., 2005).

Diagnosing: Writing

The diagnostic portion of our assessment consists of writing. Measuring writing allows the test taker to display their language ability through both their language knowledge and strategic competence. Furthermore, writing for diagnosis purposes in our context is useful because accuracy and fluency in writing comprise the bulk of learning outcomes in first-term communication classes, in which the LS classes are integrated. Following Bachman and Palmer’s (1996) model, writing involves textual knowledge, which is defined as “producing and comprehending texts that consist of two or more utterances or sentences” (p. 68). This is broken down further into knowledge of cohesion and knowledge of rhetorical or conversational organization. Writing also provides the richest pool of information for instructors to evaluate a test taker’s language ability. Although Bachman and Palmer (1996) offer a model of language abilities, Knoch (2011) points out that no specific theory of language or writing for diagnostic assessment exists; therefore, a taxonomy of available models can be best used to reflect our understanding of what is important for writing abilities for diagnostic purposes. Knoch’s (2011) taxonomy includes testable language constructs for diagnostic assessment purposes. She points out that only aspects that can be assessed through writing products should be included in these categories. Her categories consist of accuracy, fluency, complexity, mechanics, cohesion, coherence, reader/writer interaction, and content. These categories were used to develop an analytical rating scale for ESTP-O to assess test takers’ writing ability to succeed in first-term communication courses. The development of the rating scale is described further in the following section.

Self-Assessing: Language Abilities

Collecting self-reported assessment data on language abilities can add to the validity of the assessment in that correspondence between self-assessments and performance can be investigated. Li and Zhang (2021) conducted a meta-analysis of 67 studies on self-assessment and language performance. They found a moderate correlation $r = .466$ ($p < .01$) between self-assessment and language performance. They discuss the use of self-assessment and external measures of language performance, such as teacher assessments. Correlations between self-assessments and teacher assessments have been recorded in

various studies as being as high as .84 or as low as .05 (Li & Zhang, 2021). Additionally, Cox and Dewey (2021) suggest that self-assessments can be used when researchers are interested in perceptions of abilities and when low-stakes proficiency information is required. Finally, in their meta-synthesis, Zell and Krizan (2014) found moderate correlations between ability self-evaluations and performance outcomes in 22 meta-analyses. Their research shows that self-evaluations can have practical applications in that they can predict life choices (Zell & Krizan, 2014). Therefore, a test taker's choice to act on a potential recommendation to LS classes could be analyzed. Additional demographic data is also collected from the institute to categorize students into their programs and enrolment status (i.e., international or domestic students).

Methods

Research Questions

The ESTP-O is administered annually, right before the start of the fall term. The present study reports on the results from fall 2021, and addresses the following questions:

1. To what extent does receiving a recommendation for language support via the screening-diagnostic process and attending language support classes associate with students' academic achievement in their communication classes?
2. To what extent does the ESTP-O identify "at-risk" students that may need additional language support to succeed in their studies at a Canadian community college?
3. Are self-reported assessments of language abilities associated with students' performance on the diagnostic assessment or their attendance to language support classes?

ESTP-O in Its Current Form

The subsequent section outlines the testing methods in ESTP-O after its development from 2018-2020. Although the ESTP-O is not completely designed for Language for Specific Purposes (LSP), it does include more elements of workplace communication, including content from business and technical correspondence (e.g., direct messages, negative-news messages, incident reports, progress reports, job applications), rather than those more characteristic of academic writing (e.g., argumentative essays, literature reviews, reflective writing). In fact, communications instructors would like to see this assessment, especially in writing, be even more discipline specific. We focus on testing methods for vocabulary, grammar,³ and writing; they are briefly described in this order in the following paragraphs.

The screening portion of the test consisted of a vocabulary test and a grammar test. The screening portion of the test functions similarly to the DELNA in that it does not relate to follow-up pedagogical intervention, but rather merely screens for the writing diagnostic, therefore acting as a language "health check" (University of Auckland, n.d., p. 3). The current version of the vocabulary test in ESTP-O includes 20 conventional multiple-choice (MC) items, with a time limit of eight minutes. These items test written receptive vocabulary knowledge (Nation & Beglar, 2007). An example item is presented in Figure 1.

These items are developed and delivered within the institution's LMS and scored automatically.

Figure 1

Sample Vocabulary Test Items

ordeal: It was a big **ordeal**.

- a. organization
- b. trouble
- c. assignment
- d. comfort

To arrive at the current version of the vocabulary test, items were piloted and analyzed multiple times. We started with 195 vocabulary items that were derived from Nation's (2012) vocabulary size tests. Items were removed from the original pool of items if they were biased toward certain varieties of English not well known by the target audience. Words like refectory, butler, and marsupial, which are not normally used in contemporary Canadian English, were excluded. Because we aimed to measure mid-range vocabulary, some non-standard and low-frequency vocabulary (e.g., gobbet, effete, trill) from K10 or higher were also removed (Nation, 2006; Schmitt et al., 2017). We focused on mid-range frequency vocabulary (K4-K9), which we expected college-level test takers to know. Currently, 65% of the remaining items are within the mid-frequency range of the BNC-COCA (K4-K9). In 2020, using classical test theory (CTT) to measure item facility and discrimination, 69 items were analyzed and chosen to become operational. A goal of our screening-diagnostic procedure was to be sure the test was under one hour; therefore, in 2021, a reanalysis of 70 possible items was conducted to see how the test could be reduced to prevent test fatigue and demotivation. First, a CTT reliability analysis with 1,031 previous test takers was conducted. Descriptive statistics of these items are displayed in Table 1. The reliability measure was $\alpha = .93$.

Table 1

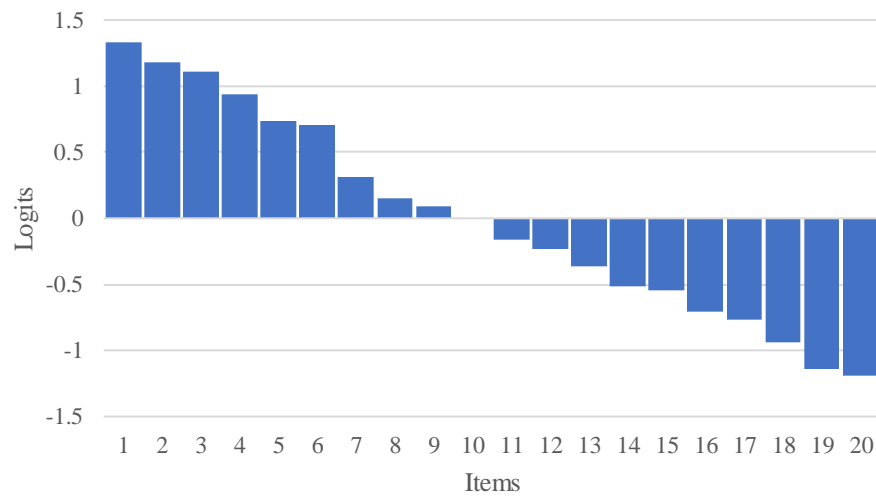
Descriptive Statistics for 2020 Vocabulary Items (k=70)

<i>N</i>	1,031
<i>M</i>	59.8
Median	63
<i>SD</i>	9.52
Kurtosis	2.92
Skewness	-1.71

The Spearman-Brown prophecy formula was then used to estimate the reliability of a shorter test. The formula estimated that a test of 20 items would still have a reliability of approximately 0.78, which is acceptable for a low-stakes test (Green, 2013). Second, a 2-parameter (PL) Rasch analysis was conducted with the same 70 items. Twenty items within the ideal infit and outfit range of 0.5 and 1.5 were retained to fit our model (Bond & Fox, 2015; Green, 2013).

Figure 2 shows that the item difficulty range of these 20 remaining items fell between 1.33 and -1.14, measured in logits.

Figure 2
Vocabulary Item Difficulty in Logits



The score table in Table 2 shows what test takers were predicted to score based on their person ability measured by theta.

Table 2
Vocabulary Score Table

Score	Theta	SE
0	-4.45	1.85
1	-3.19	1.04
2	-2.42	0.76
3	-1.93	0.65
4	-1.56	0.58
5	-1.24	0.54
6	-0.97	0.51
7	-0.71	0.5
8	-0.47	0.49
9	-0.24	0.48
10	-0.01	0.48
11	0.22	0.48
12	0.46	0.49
13	0.70	0.5
14	0.96	0.52
15	1.24	0.55
16	1.56	0.59
17	1.94	0.65
18	2.43	0.77
19	3.21	1.04
20	4.47	1.85

Table 2 indicates that a test taker with a theta of almost 0 would have a chance of scoring 50% on the test.

The second part of the screening portion of the test consists of a tailored cloze test designed to test implicit grammar knowledge (Brown, 2013; Ellis, 2005). Grammar was tested because grammatical accuracy is considered important for instructors in business and technical communication classes, especially in fields such as healthcare, engineering, and business where grammatical errors can lead to miscommunication which can result in health, safety, or financial risks. Knoch and Macqueen (2020) underscore that language assessment for professional purposes is part of overall risk management for professionals in high-risk jobs where communication breakdowns can have serious consequences. The test sought to measure test takers' morphosyntactic knowledge (Di Biase & Kawaguchi, 2013; Pienemann, 1998) and knowledge of cohesion (Halliday & Hasan, 1976/2013; Purpura, 2004). This included testing for the understanding of word order, subject-verb agreement, affixes, verb tense and aspect, articles, reference and substitution, lexicon cohesion, as well as conjunctions and text connectives. Descriptive statistics of the grammar test based on an analysis of 404 test takers from fall 2020 are displayed in Table 3.

Table 3

Descriptive Statistics of Fall 2020 Grammar Test (k=30)

<i>N</i>	404
<i>M</i>	16.9
Median	18
<i>SD</i>	6.32
Kurtosis	-0.37
Skewness	-0.57

A CTT analysis after the 2020 administration suggested that seven items had poor to borderline item-rest correlation ($< .30$) and were removed. Three additional items were removed because of multiple possible correct answers and repetition of answers. Therefore, for 2021, 20 grammar items remained which had an internal reliability of $\alpha = 0.88$. The grammar test also had a secondary function as a “check and balance” in case students looked up vocabulary words during the first part of the screening. As the test is not invigilated, the grammar portion was timed at five minutes to reduce the chance of sharing answers with other test takers.

If students do not meet the cut score in the screening portion of the test, they are navigated to a 20-minute writing diagnostic within the LMS. The writing diagnostic aims to measure students' ability to interpret and create texts by using their language knowledge and strategic competence (Bachman & Palmer, 2010). Writing, according to stakeholders (i.e., communication and LS instructors) who were consulted as part of the overall development of test tasks, was the ability for which at-risk students needed the most support. Therefore, with the help of instructors, these performance tasks were developed with the TLU domain in mind and revolved around writing school or work-related emails. Writing clear and concise workplace emails is a major focus in first-term business and technical communication classes. Email writing also plays a large role in workplace communication in business and technical settings (Darics & Koller, 2018; Ewald, 2020) and “can have an impact not only on learning, but on social relations in both school and the workplace” (Dimova et al., 2020, p. 81). Email writing in higher education is also often

overlooked, as students can spend one to two hours per day on average reading and writing emails, according to Dimova et al. (2020).

The prompts were revised from communication instructors' retired writing tasks for diagnostic or exam purposes. Therefore, these were authentic tasks that students could experience in first-term communication classes. First, we reviewed these tasks for fairness and ensured the readability levels of the prompts were appropriate for a college-level target audience (Flesch Kincaid grade level range: Min = 8.44, Max = 10.97). Second, we repackaged the tasks so that they were uniform and met best practice criteria for writing task development (Douglas, 2000). For instance, the tasks start with a title that indicates the general topic and a paragraph that describes the communicative situation (Douglas, 2000); this is a maximum of 60 words long. The paragraph is followed by instructions (maximum 20 words) that include three to four bullet points about what should be included in the response, how many words the email should be, as well as the task time. Finally, four bullet points explain how the response will be evaluated. Communication instructors can choose from five different prompts for their specific program. An example writing task is found in Figure 3.

Figure 3

Sample Diagnostic Writing Task

1. Requesting Career Program Support

Situation: BCIT Career Edge is a program that offers job search support for BCIT students. One of their services is career coaching. Career coaches can provide guidance on searching for jobs, preparing application documents, identifying companies that are hiring, and building a professional network. You are interested in getting support from this program.

Instructions: Write an email to the program administrator (Alex Lee).

- Briefly introduce yourself and your program of study.
- Describe the type of job you are looking for.
- Describe your relevant, school, volunteer, and/or work experience.
- Explain the type of support you are looking for and why.

Write a 250-400 word email (about 10-20 sentences). You have 20 minutes for this task.

Your response will be evaluated on your ability to:

- Sufficiently address and organize all parts of the task and provide details.
- Use a wide range of vocabulary with accuracy and few spelling errors.
- Use a variety of sentence structures both accurately and effectively.
- Create a clear message that logically connects ideas.

To measure writing, a rating scale based on Kuiken and Vedder (2016) was originally developed (see Devos, 2019). This rating scale contained six criteria, including content, persuasion, genre, tone, comprehensibility, and cohesion across six bands. To make the marking procedure of the diagnostic more feasible for instructors and create positive washback in LS classes, a new online rating scale and writing prompts were developed.

Using Knoch's (2011) taxonomy of testable criteria for diagnostic assessment, writing task criteria were developed from the CLB (2012) and input from instructors in first-term communication courses. The current analytic rating scale contains four criteria: (1) content/task, (2) vocabulary range and accuracy, (3) grammatical range and accuracy,

and (4) coherence and cohesion. Each criterion is described on three bands: (3) program ready, (2) borderline, and (1) needs language support. Writing samples are scored by LS instructors on a scale of 4 to 12. Test takers with a score of 6 or below receive a recommendation for LS classes, students with scores of 7 to 9 are informed of a number of English language support services at the institute (e.g., online learning materials, writing centre, peer conversation group), and students who score between 10 and 12 receive no notification. Each band includes descriptors that are based on CLB as well as communication course outcomes so that LS instructors (who are also trained and experienced EAL professionals) can associate standardized evaluation criteria and their programs' writing goals with the rating scale. We have since equated these scores to the CLB band descriptors for writing by developing performance level descriptors (PLDs) (Mehrens & Cizek, 2012). Students who score in the lowest band are considered to be at a CLB 4-6 range in their writing abilities, borderline between CLB 7-8, and program-ready at CLB 9 or above.

The online rating scale is embedded in the LMS so that instructors can view writing samples, score them with the rubric, and provide written feedback that students can read afterwards. Furthermore, the vocabulary and grammar bands include descriptors that are based on ongoing research on grammar testing in college diploma programs. For example, vocabulary accuracy includes focusing on word forms as well as count and non-count nouns, whereas grammatical accuracy focuses on verb tense and aspect, subject-verb agreement, articles, prepositions, and reference.

Instructors in 16 of the 23 programs opted to use the screening test plus writing diagnostic option offered by the test development team. The remaining programs opted for only the screening portion of the test to corroborate their own writing diagnostics.

Finally, a short six-item survey accompanies the test each term. The survey is embedded in the LMS via an online survey platform and is completed before the start of the test. This survey allows us to compare how students perform relative to their gender, age, and the language they are most confident using. They are also asked to assess their own language abilities on a three-level scale (*basic, intermediate, advanced*) in the four skill areas, reading, writing, speaking, and listening.

Participants

After removing invalid and incomplete entries, as well as removing students who took the test but dropped the program before it began, we were left with 1,388 valid test and survey entries. Test taker demographics are displayed by school in Table 4.

Table 4
Test Taker Demographics

	<i>N</i>	<i>Gender</i>		<i>Age</i>		<i>Enrolment</i>	
		<i>M</i>	<i>F</i>	<i>Mean</i>	<i>SD</i>	<i>International</i>	<i>Domestic</i>
School of Business and Media	782	419	363	22.1	5.4	215	563
School of Energy	270	238	32	21.4	4.8	49	220
School of Construction and the Environment	255	153	102	22.3	5.7	36	219
School of Health Sciences	59	30	29	25.9	8.1	16	43
School of Computing and Academic Studies	22	11	11	24.5	5.8	4	18

Overall, about 60% of the participants were male ($n = 851$) and 40% female ($n = 537$). The overall age of the test takers was 22.2 years old ($SD = 5.6$). The School of Energy, which includes engineering programs, had the highest male-to-female difference, with 66% more males than females. Overall, domestic students made up 77% of the test takers and international students comprised 23%. The School of Business and Media had the highest percentage of international enrolments with 27%. The School of Construction and the Environment had the lowest percentage with 14% international enrolments. Only about 9% of all the test takers had been through one of the institution's pre-entry programs. All other test takers had met the institution's English language prerequisites through prior English-speaking education, pathway schools, or standardized English language proficiency assessments (e.g., IELTS, TOEFL iBT, CAEL, DET, etc.).

Data analysis procedures

Test and survey data were collected in the secure LMS and online survey platform and preprocessed using MS Excel 2016 before being analyzed using SPSS (Version 27) and RStudio (Version 1.4.1106). Descriptive statistics were used to answer the research question about to what extent the ESTP-O identifies test takers for LS classes. In addition, to validate the use of the vocabulary test as a screening mechanism for releasing the writing diagnostic portion of the test, correlation analysis and multilevel regression analysis were conducted to test the null hypothesis that there is no relationship between test takers' scores on the vocabulary test and their academic achievement in communication classes. For the first research question, we tested the null hypothesis that there is no association between receiving a recommendation and attending LS classes and academic achievement in communication classes. To do this, a multilevel regression model was built to control for age, gender, and the effect of the instructor. Finally, to answer the third research question we used the chi-square (X^2) test for independence and Fisher-Freeman-Halton's exact test to test associations between the self-assessments of language abilities and participants' recommendation to LS and their attendance. Here, the null hypothesis was that there would be no association between participants' self-reporting of language abilities and their recommendation to LS classes or their participation in these classes.

Results and Discussion

Language Support and Academic Achievement in Communication Classes

Our first research question was: To what extent does receiving a recommendation for LS via the screening-diagnostic process and attending these classes associate with students' academic achievement in their communication classes? We pursued this question by testing the association between students receiving a recommendation at the start of the term via a screening-diagnostic procedure and attending classes (or not) and their final grades in communication classes. Students who received a recommendation to attend LS either via the development team's screening screening-diagnostic procedure or via the screening plus instructor's own diagnostic were included in this analysis.

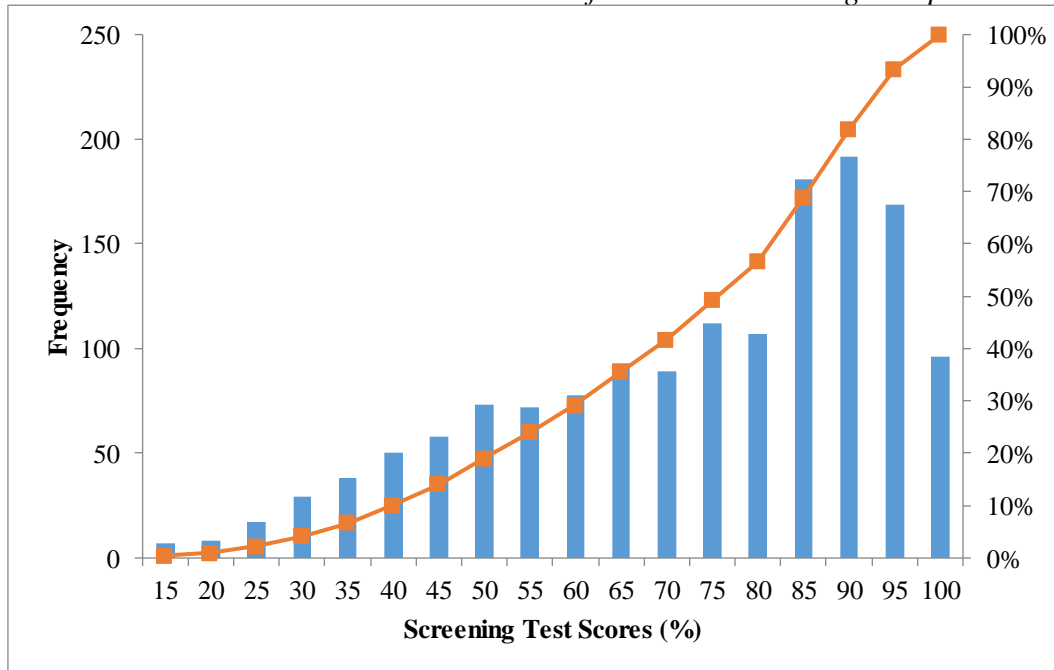
Screening Component

Step one in the procedure included students taking the screening portion of the test, which consists of the total score of vocabulary and grammar. To reduce the risk of missing students who may require support (false negatives), a cut score of 60% on the screening component was decided on prior to administration.

The score distribution of the screening component is presented in Figure 4. In total, 430 test takers (29.3%) fell below the cut score.

Figure 4

Score Distribution and Cumulative Percent of Scores in Screening Component

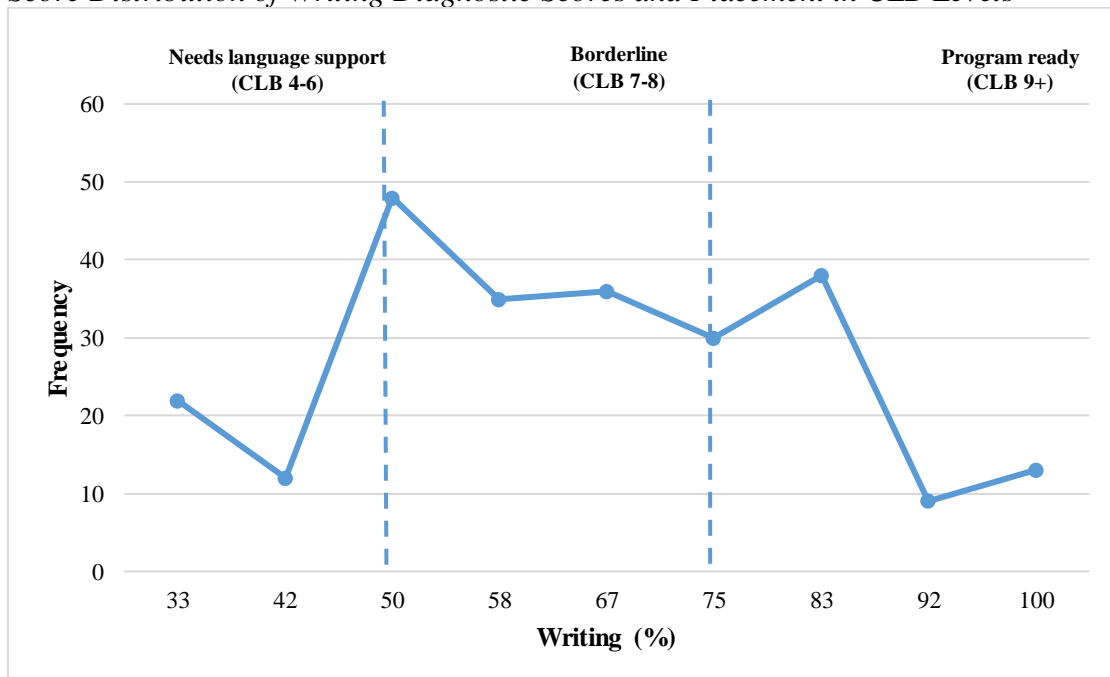


Writing Diagnostic

The writing diagnostic is step two of the procedure and scored using the online analytic rating scale by LS instructors who are trained EAL instructors. As only 16 programs opted for the full screening-diagnostic procedure, we analyzed a subset ($n = 722$) of test takers who participated in the full procedure ($M = 429$, $F = 258$, mean age = 22.9, $SD = 5.9$). From this subset, 243 (34%) fell below the cut score and were navigated to the writing task. Figure 5 presents the distribution of their writing diagnostic scores (converted to percentages) and an indication of the CLB bands.

Figure 5

Score Distribution of Writing Diagnostic Scores and Placement in CLB Levels



As Figure 5 shows, the writing diagnostic places test-takers into three bands, i.e., needs language support, borderline, and program ready. From these 243 written responses, 82 (34%) were recommended to LS classes. Twenty-five percent of the students were considered program ready, while 42% of the students were identified as borderline and either received a recommendation to attend LS classes or additional information about other English support or central writing services at the institution. Based on the above results and recommendations based on instructors' own writing assignments, a total of 262 students were recommended at the start of the term for LS classes.

To answer our first research question about the extent to which receiving a recommendation for LS through the two-step process and attending LS classes is associated with final communications grades, we took a sample of 1,268 students across 74 communication classes, approximately 60% males and 40% females. Their ages ranged from 17-52 years old, with a mean age of 22.1 years old. We controlled for age and gender, which were significant predictors of grades ($p < .001$). The interclass correlation coefficient (ICC) indicated that 21% of the variation in communication grades was due to the

instructor, so we controlled for the effect of the instructor in the model. Communication class grades were then analyzed following the above three categories and the results are presented in Table 5.

Table 5

Mean Communication Grades by Language Support Group

Groups	<i>M</i>	<i>SE</i>	<i>df</i>	95% Confidence Interval	
				<i>Lower</i>	<i>Upper</i>
Not recommended	78.98	0.69	75.71	77.6	80.36
Recommended and attended	69.89	1	54.24	67.89	71.9
Recommended and did not attend	64.17	1.93	48.73	60.3	68.04

Note. Estimated means are estimated keeping constant other effects in the model to the mean

Table 5 shows that when all else is held constant, students who received a recommendation and attended LS classes had mean communication grades of almost 70%, whereas students who received a recommendation and did not attend had mean communication grades of about 64%. These results confirm previous research findings attesting to significant improvement following post-admission language intervention (Donohue, & Erling, 2012; Skinner & Mort, 2009; Urmston et al., 2016).

Vocabulary May Predict Students' Performance in Communication Classes

Our second research question was: To what extent does the ESTP-O identify “at-risk” students that may need additional language support to succeed in their studies at a Canadian community college? This was answered by testing the hypothesis of whether there was a relationship between students' communication grades and their vocabulary scores on ESTP-O, which was done at the end of the fall term. If the vocabulary test had predictive validity, it may add to its overall validity and confirm its use as a screening mechanism for recommending students to LS classes. Academic achievement in communication classes was also chosen because the goal of language support is to help students in meeting the language-based outcomes in these courses. Taking a sample of 1,266 students across 74 communication classes (60% male, 40% female; mean age = 22), we conducted a correlation analysis and a multilevel regression analysis. The results of the correlation analysis showed that the relationship between communication grades and vocabulary scores was moderately positive ($r = 0.36, p < .001$). A multilevel regression model was built to adjust for the effect of the instructor, age, and gender, which accounted for 37% of the variance in the model (R^2 conditional = 0.37). The results of the regression model are presented in Table 6.

Table 6
Fixed Effects Parameter Estimates

<i>Names</i>	<i>Effect</i>	<i>Estimate</i>	<i>SE</i>	95% Confidence Interval		<i>Df</i>	<i>T</i>	<i>p</i>
				<i>Lower</i>	<i>Upper</i>			
(Intercept)	(Intercept)	74.71	0.73	73.28	76.14	88.63	102.58	< .001
Vocabulary	Vocabulary	0.25	0.02	0.20	0.29	63.82	11.04	< .001
Gender	Gender	3.44	0.66	2.14	4.73	1250.23	5.19	< .001
Age	Age	0.37	0.06	0.26	0.49	1239.32	6.51	< .001

As can be seen from Table 6, the results show a significant non-zero mean in communication grades and a significant slope in vocabulary scores. The regression coefficient indicates, therefore, that, on average, a 1 percentage point increase in vocabulary test scores is associated with a 0.25 percentage point increase in communication grades, when all else is held constant. This model suggests that vocabulary test scores can be significant predictors of students' communication grades, after controlling for gender, age, and the effects of the instructor. These results answer the second research question examined in this study and confirm previous research establishing a predictive relationship between vocabulary knowledge and overall language ability (Meara, 1996; Miralpeix & Muñoz, 2018) as well as academic achievement (Coxhead, 2000; Coxhead & Nation, 2001).

Association Between Students' Self-Reported Writing Abilities, Performance, and Attendance

The third research question examined in this study is: Are self-reported assessments of language abilities associated with students' performance on the diagnostic assessment or their attendance to LS classes? To answer this question, we looked at the associations between both students' assessment of their writing abilities and their performance on the writing diagnostics as well as their responses and subsequent attendance to LS classes.

Respondents were asked to self-assess their English abilities in reading, writing, listening, and speaking on a scale of *basic*, *intermediate*, or *advanced*. Table 7 shows the results of the self-report survey in percentages. Of the four skills, 9% of students rated their writing abilities as *basic*, with 48% rating their writing as *intermediate*, and 43% as *advanced*. The other three skills had most students reporting them as advanced. The highest was listening, with 64% of students reporting this ability as *advanced*.

Table 7
Self-Assessment of Language Abilities

	<i>N</i>	<i>Basic</i>	<i>Intermediate</i>	<i>Advanced</i>
Reading	1,415	5%	37%	58%
Writing	1,419	9%	48%	43%
Listening	1,416	3%	32%	64%
Speaking	1,416	6%	37%	57%

As participants were recommended to LS via the writing diagnostic, the subsequent analyses focused on 668 students ($M = 398$, $F = 239$, mean age = 22.9, $SD = 5.9$) who participated in the full screening and diagnostic option and included a self-report of their writing abilities; 23% were international students, while 73% were domestic students. The analyses concentrated on the evaluation by instructors of written responses, which resulted in a recommendation to LS classes or not. The first analysis tested the hypothesis of an association between self-assessment of writing skills and the bivariate of being recommended for language support (1) or not (0). Using the chi-square (X^2) test for independence, we compared these two variables with participants' assessment of their ability in writing (*basic*, *intermediate*, or *advanced*). The test suggested that we reject the null hypothesis and can accept the alternative hypothesis that there is an association between self-reported writing abilities and a referral to LS (see Table 8).

Table 8
Contingency Table of Self-Assessment of Writing Abilities and Instructors' Recommendation to LS

	Self-Report of Writing Abilities		
	<i>Basic</i>	<i>Intermediate</i>	<i>Advanced</i>
Recommended to LS			
No	39	260	288
Yes	15	62	4

Note. Pearson Chi-Square = 59.48, $df = 2$, $p < .05$

Zell and Krizan (2014) suggest that self-reporting on abilities may predict life choices, so to test whether there was an association between self-assessed writing abilities and attendance to LS, we ran a second analysis. For this, we used the Fisher-Freeman-Halton exact test for 81 participants ($M = 50$, $F = 25$, mean age = 22.5, $SD = 6.2$) who had valid self-assessment responses on their writing abilities and the bivariate of attending LS classes (1) or not (0). The null hypothesis was that there would be no association between students' self-reporting of writing abilities and attendance to LS classes during the semester. The result showed that there was not a significant association between these two variables (two-tailed $p = .101$). In sum, the self-assessment of writing abilities was associated with students' receiving a recommendation for LS but was not associated with their attendance.

Conclusion

Internationalization and immigration have created an influx of students in college diploma programs with diverse language backgrounds. Some students enter SCTE technical and business college diploma programs with language abilities that do not meet program expectations and, therefore, academic language support is required. This paper reports on the development and administration of a locally developed post-admission language diagnostic assessment (i.e., ESTP-O) to recommend at-risk students to integrated, non-credit-bearing LS classes in college diploma programs. We present some validity evidence for the use of this screening-diagnostic procedure to identify and recommend students to weekly classes that aim to mitigate their risk of failing in communication classes in their programs.

The goal of the ESTP-O was to create a local test that is fit for the purpose of offering diagnostic information on students' writing skills followed by pedagogical support. To accommodate the move to online learning in 2020 and 2021, this assessment was delivered online and remotely via the institution's LMS. In the fall of 2021, 1,388 students from 23 different technical and business diploma programs took the ESTP-O. About 30% of the test takers fell below a pre-determined cut score on the screening test. An analysis of the vocabulary test as an adequate method for determining who should take the writing diagnostic suggested that it has significant predictive validity. From the writing portion of the test, 82 (11%) of the students were recommended for LS classes as they were considered at risk of failing their communication courses. The validity of this recommendation process and the pedagogical intervention was determined by analyzing the academic achievement of 1,268 students in 74 communication classes. It was discovered that students who received a recommendation and attended LS classes had grades that were 6% higher than students who were recommended but did not attend. Finally, we also analyzed students' self-reported language abilities for associations with a recommendation to LS classes by trained instructors and their subsequent attendance to these classes. Although there was no association with attendance, students' self-reported writing abilities did have an association with the instructor's evaluation of their written responses.

Naturally, the validation of ESTP-O is an ongoing and iterative process. As Zumbo (2007) aptly states, "Measurement or test score validation is an ongoing process wherein one provides evidence to support the appropriateness, meaningfulness, and usefulness of the specific inferences made from scores about individuals from a given sample and in a given context" (p. 48). Therefore, our test has multiple limitations. First, we only look at the vocabulary from one dimension, written receptive knowledge. Research suggests that vocabulary breadth (Meara, 1996; Miralpeix & Muñoz, 2018; Nation & Beglar, 2007) and depth (Nizonkiza, 2011) are important in measuring vocabulary as a predictor of language growth and proficiency. Second, we are still uncovering the validity of grammar as a screening mechanism. That is, is grammar as a construct valid in predicting students' academic achievement in communication courses? A new grammar test is currently being developed and piloted by the test development team to determine which grammar errors might trigger content instructors' evaluation of students' writing and, thus, determine them as at-risk. Some literature and research exist in this area (Biber et al., 2011; di Gennaro, 2016; Kyle, 2018; Lahuerta, 2018; Lastres-López & Manalastas, 2017; MacDonald, 2016; Shapiro et al., 2014), but further research into the importance of these grammar structures in the local context is required. Third, rater training on rater-mediated assessments is

important for scale consistency (Dimova et al., 2020). Although LS instructors were involved in the development of the rating scale and received some orientation via a training video, we were unable to train the instructors extensively on the scale before test administration. Related to this, we could not determine the inter-rater reliability of the written responses because it is not feasible to have two raters mark writing samples. Finally, the use of the LMS as an online, remote testing platform may introduce construct irrelevant variance (Chapelle & Voss, 2017; Messick, 1996). However, a recent report by Zumbo (2021) suggests that performance on computer-based, invigilated language tests taken at a testing centre was the same as those taken remotely, online at home. Nonetheless, a more suitable platform to reduce the potential interference of technology on measuring language constructs is desirable. Finally, our test and results are confined to our local context and based on a small sample of first-term students in college diploma programs. Therefore, the results cannot be generalized to other populations or institutions. However, future research can investigate the use of similar tools to evaluate and recommend students for language support post-admissions.

In sum, the development of a local screening-diagnostic process is possible for evaluating and recommending at-risk students to embedded LS classes in college diploma programs. The ESTP-O was able to identify at-risk students and offer them pedagogical support that helped them academically, supporting Read's (2015a, 2016) consequential validity argument for post-admission assessment. While this process is resource intensive and requires an iterative development process to assess and investigate the validity of the test, we hope the results give us a solid foundation for continuing to develop a valid assessment that offers diagnostic information on students' strengths and weaknesses so that they can succeed in their programs and graduate to find meaningful work opportunities.

Correspondence:

Correspondence should be addressed to: Nathan Devos

Email: ndevos1@bcit.ca

Notes:

¹ See <https://arts.unimelb.edu.au/language-testing-research-centre/language-assessments/ready-made-tests>. Retrieved March 16, 2023.

² English Studies 12 is the province of BC's English language arts course required for high school graduation. An equivalent course is English First Peoples 12 or English Literature 12. This course ends with a provincial literacy assessment.

³ We present the grammar test in its state in fall 2021; however, it has since been grandfathered and a new grammar test is current being developed and piloted.

References

- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. Bloomsbury Academic.
<http://dx.doi.org/10.5040/9781474212151>
- Alderson, J. C. (2007). The challenges of diagnostic testing: Do we know what we are measuring? In J. Fox, M. Wesche, D. Bayliss, L. Cheng, & C. E. Turner (Eds.), *Language testing reconsidered* (pp. 21-30). University of Ottawa Press.
- Alderson, J. C., Brunfaut, T., & Harding, L. (2015). Towards a theory of diagnosis in second and foreign language assessment: Insights from professional practice across diverse fields. *Applied Linguistics*, 36(2), 236–260.
<https://doi.org/10.1093/applin/amt046>
- Alderson, J. C., & Huhta, A. (2005). The development of a suite of computer-based diagnostic tests based on the Common European Framework. *Language Testing*, 22(3), 301–320. <https://doi.org/10.1191/0265532205lt310oa>
- Arkoudis, S., Baik, C., & Richardson, S. (2012). *English language standards in higher education: From entry to exit*. ACER Press.
<https://doi.org/10.1080/07294360.2013.783955>
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessment and justifying their use in the real world*. Oxford University Press.
- Beynen, T. (2020) Metaphor comprehension and engineering texts: Implications for English for Academic Purposes (EAP) and first-year university student success, *TESL Journal Canada/Revue TESL Du Canada*, 37(1), 22-50.
<https://doi.org/10.18806/tesl.v37i1.1332>
- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45(1), 5–35. <http://www.jstor.org/stable/41307614>
- Bonanno, H., & Jones, J. (2007). *The MASUS procedure: Measuring the academic skills of university students: A diagnostic assessment*. University of Sydney, Learning Centre.
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. (3rd ed.). Routledge.
<https://doi.org/10.4324/9781315814698>
- Brown, J. D. (2013). My twenty-five years of cloze testing research: So what? *International Journal of Language Studies*, 7(1), 1-32.
<http://www.ijls.net/sample/71-1.pdf>
- CLB. (2012, October). *Canadian language benchmarks: English as a Second Language for Adults*. Centre for Canadian Language Benchmarks.
<https://www.canada.ca/content/dam/ircc/migration/ircc/english/pdf/pub/language-benchmarks.pdf>
- Celce-Murcia, M. (2007). Rethinking the role of communicative competence in language teaching intercultural language use and language learning. In E.A. Soler, & M.P. Jordà (Eds.), *Intercultural language use and language learning* (pp. 41–57). Springer.

- Chapelle, C.A., Voss, E. (2017). Utilizing technology in language assessment. In E. Shohamy, I. Or, & S. May (Eds.), *Language testing and assessment. Encyclopedia of language and education* (pp. 149-161). Springer.
- Council of Europe. (2020). *Common European Framework of References for Languages: Learning, teaching, assessment- Companion volume*. Council of Europe Publishing. www.coe.int/lang-cefr.
- Cox, T. L., & Dewey, D. P. (2021). Measuring language development through self-assessment. In P. Winke, & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 382-390). Routledge.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238. <https://doi.org/10.2307/3587951>
- Coxhead, A., & Nation, P. (2001). The specialised vocabulary of English for academic purposes. In J. Flowerdew, & M. Peacock (Eds.), *Research perspectives on English for academic purposes*, 252–267. <https://doi.org/10.1017/CBO9781139524766.020>
- Crossman, E., Choi, Y., Lu, Y., & Hou, F. (2022, March 23). *International students as a source of labour supply: A summary of recent trends*. Statistics Canada. <https://www150.statcan.gc.ca/n1/pub/36-28-0001/2022003/article/00001-eng.htm>
- Darics, E., & Koller, V. (2018). *Language in business, language at work*. Bloomsbury.
- Devos, N. J. (2019). Comparing first-term students' English language proficiency at a Canadian polytechnic institute. *BC TEAL Journal*, 4(1), 53-83. <https://doi.org/10.14288/bctj.v4i1.335>
- Devos, N. J. (2022, November 7). *BCIT provides students with free English language support services*. BCIT News. <https://commons.bcit.ca/news/2022/11/english-language-support-services/>
- Di Biase, B., & Kawaguchi, S. (2013). Processability theory. In P. Robinson (Ed.), *The Routledge encyclopedia of second language acquisition* (pp. 512-518). Routledge.
- di Gennaro, K. (2016). Searching for differences and discovering similarities: Why international and resident second-language learners' grammatical errors cannot serve as a proxy for placement into writing courses. *Assessing Writing*, 29, 1-14. <https://doi.org/10.1016/j.asw.2016.05.001>
- Dimova, S., Yan, X., & Ginther, A. (2020). *Local language testing: Design, implementation, and development*. Routledge.
- Donohue, J. P., & Erling, E. J. (2012). Investigating the relationship between the use of English for academic purposes and academic attainment. *Journal of English for Academic Purposes*, 11(3), 210-219. <https://doi.org/10.1016/j.jeap.2012.04.003>
- Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge University Press.
- Dunworth, K. (2009). An investigation into post-entry English language assessment in Australian universities. *Journal of Academic Language and Learning*, 3(1), A1-A13. <https://journal.aall.org.au/index.php/jall/article/view/67>
- Dyson, B. P. (2009). Understanding trajectories of academic literacy: How could this improve diagnostic assessment? *Journal of Academic Language and Learning*, 3(1), A52-A69. <https://journal.aall.org.au/index.php/jall/article/view/99>
- Edwards, E., Goldsmith, R., Havery, C., & James, N. (2021). An institution-wide strategy for ongoing, embedded academic language development: Design, implementation and analysis. *Journal of Academic Language and Learning*, 15(1), 53-71.

- Elder, C., & Erlam, R. (2001). *Development and validation of the Diagnostic English Language Needs Assessment (DELNA): Final report*. The University of Auckland, Department of Applied Language Studies and Linguistics.
<https://cdn.auckland.ac.nz/assets/delna/delna/documents/elder-erlam-2001-report.pdf>
- Elder, C., & Read, J. (2015). Post-entry language assessments in Australia. In J. Read (Ed.), *Assessing English proficiency for university study* (pp. 25-46). Palgrave Macmillan.
- Elder, C., & von Randow, J. (2008). Exploring the utility of a web-based English language screening tool. *Language Assessment Quarterly*, 5(3), 173–194.
<https://doi.org/10.1080/15434300802229334>
- Ellis, R. (2005). Measuring implicit and explicit knowledge of second language: A psychometric study. *Studies in Second Language Acquisition*, 27, 141-172.
<https://doi.org/10.1017/S0272263105050096>
- Erlam, R., & Botelho de Magalhães, M. (2021). Post-entry English language assessment at the University of Auckland: Ongoing validation of DELNA. *New Zealand Studies in Applied Linguistics*, 27(1), 32–49.
- Erling, E. J., & Richardson, J. T. E. (2010). Measuring the academic skills of university students: Evaluation of a diagnostic procedure. *Assessing Writing*, 15(3), 177–193.
<https://doi.org/10.1016/j.asw.2010.08.002>
- Ewald, T. (2020). *Writing in the technical fields: A practical guide*. (3rd ed). Oxford University Press.
- Fox, J. (2005). Test decisions over time: Tracking validity. *Language Testing*, 21(4), 437–465. <https://doi.org/10.1191/0265532204lt292oa>
- Fox, J. (2015). Trends and issues in language assessment in Canada: A consideration of context. *Language Assessment Quarterly*, 12(1), 1–9.
<https://doi.org/10.1080/15434303.2014.999921>
- Fox, J., Haggerty, J., & Artemeva, N. (2016). Mitigating risk: The impact of a diagnostic assessment procedure on the first-year experience in engineering. In J. Read (Ed.), *Post-admission language assessment of university students* (pp. 43–65). Springer International Publishing. https://doi.org/10.1007/978-3-319-39192-2_3
- Fox, J., von Randow, J., & Volkov, A. (2016). Identifying students at-risk through post-entry diagnostic assessment: An Australasian approach takes root in a Canadian university. In V. Aryadoust & J. Fox (Eds.), *Trends in language assessment research & practice: The view from the Middle East and Pacific Rim* (pp. 265-285). Cambridge Scholars Publishing.
- 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>
- Fulcher, G. (1997). An English language placement test: Issues in reliability and validity. *Language Testing*, 14, 113–139. <https://doi.org/10.1177/026553229701400201>
- Green, R. (2013). *Statistical analysis for language testers*. Palgrave Macmillan.
- Halliday, M. A. K., & Hassan, R. (2013). *Cohesion in English*. Routledge. (Original work published in 1976)
- Heeren, J., Speelman, D., & De Wachter, L. (2021). A practical academic reading and vocabulary screening test as a predictor of achievement in first-year university students: Implications for test purpose and use. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1458-1473.
<https://doi.org/10.1080/13670050.2019.1709411>

- Hirsch, R. (2020). An interview with Dr. John Read. *Language Assessment Quarterly*, 17(2), 204–215. <https://doi.org/10.1080/15434303.2020.1730842>
- Holder, G. M., Jones, J., Robinson, R. A., & Krass, I. (1999). Academic literacy skills and progression rates amongst pharmacy students. *Higher Education Research and Development*, 18(1), 19–30. <https://doi.org/10.1080/0729436990180103>
- Jang, E. E. (2009). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for fusion model application to language assessment. *Language Testing*, 26(1), 31–73. <https://doi.org/10.1177/0265532208097336>
- Knoch, U. (2011). Rating scales for diagnostic assessment of writing: What should they look like and where should the criteria come from? *Assessing Writing*, 16, 81–96. <https://doi.org/10.1016/j.asw.2011.02.003>
- Knoch, U., & Macqueen, S. (2020). *Assessing English for professional purposes*. Routledge. <https://doi.org/10.4324/9780429340383>
- 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>
- Kyle, K. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *Modern Language Journal*, 102(2), 333–349. <https://doi.org/10.1111/modl.12468>
- Lahuerta, A. C. (2018). Study of accuracy and grammatical complexity in EFL writing. *International Journal of English Studies*, 18(1), 71–89. <https://doi.org/10.6018/ijes/2018/1/258971>
- Lastres-López, C., & Manalastas, G. (2017). Errors in L1 and L2 university students' writing in English: Grammar, spelling and punctuation. *Revista Electrónica De Lingüística Aplicada*, 16(1), 118–134. <http://www.aesla.org.es/ojs/index.php/RAEL/article/view/325>
- Laufer, B., & Levitzky-Aviad, T. (2018). Loanword proportion in vocabulary size tests: Does it make a difference? *International Journal of Applied Linguistics*, 169(1), 95–114. <https://doi.org/10.1075/itl.00008.lau>
- Li, M., & Zhang, X. (2021). A meta-analysis of self-assessment and language performance in language testing and assessment. *Language Testing*, 38(2), 189–218. <https://doi.org/10.1177/0265532220932481>
- MacDonald, P. (2016). “We all make mistakes!”. Analysing an error-coded corpus of Spanish university students' written English. *Complutense Journal of English Studies*, 24, 13–129. <http://dx.doi.org/10.5209/CJES.53273>
- Mehrens, W. A., & Cizek, G. J. (2012). Standard setting for decision making: Classifications, consequences, and the common good. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations*. (2nd ed.). (pp. 33–46). Routledge. <https://doi.org/10.4324/9780203848203>
- Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Competence and performance in language learning* (pp. 35–53). Cambridge University Press.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–156. <https://doi.org/10.1177/026553229601300302>
- Miralpeix, I., & Muñoz, C. (2018). Receptive vocabulary size and its relationship to EFL language skills. *International Review of Applied Linguistics in Language Teaching*, 56(1), 1–24. <https://doi.org/10.1515/iral-2017-0016>

- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 63(1), 59-82. <https://doi.org/10.3138/cmlr.63.1.59>
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13. http://www.jalt-publications.org/archive/tlt/2007/07_2007TLT.pdf
- Nation, I. S. P. (2012, October 23). *The vocabulary size test*. Victoria University of Wellington, School of Linguistics and Applied Linguistics. <https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Vocabulary-Size-Test-information-and-specifications.pdf>
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/9781009093873>
- Nizonkiza, D. (2011). The relationship between lexical competence, collocational competence, and second language proficiency. *English Text Construction*, 4(1), 113-146. <https://doi.org/10.1075/etc.4.1.06niz>
- Norris, J. M. (2012). Purposeful language assessment: Selecting the right alternative test. *English Teaching Forum*, 3, 41-45. <https://files.eric.ed.gov/fulltext/EJ997530.pdf>
- Palmer, L., Levett-Jones, T., & Smith, R. (2018). First year students' perceptions of academic literacies preparedness and embedded diagnostic assessment. *Student Success*, 9(2), 49-61. <https://doi.org/10.5204/ssj.v9i2.417>
- Palmer, L., Levett-Jones, T., Smith, R., & McMillan, M. (2014). Academic literacy diagnostic assessment in the first semester of first year at university. *The International Journal of the First Year in Higher Education*, 5(1), 67-78. <https://doi.org/10.5204/intjfyhe.v5i1.201>
- Paton, M. J. (2007). Why international students are at greater risk of failure: An inconvenient truth. *International Journal of Diversity*, 6(6), 101-111. <https://doi.org/10.18848/1447-9532/CGP/v06i06/39297>
- Pienemann, M. (1998). *Language processing and second language development: Processability theory*. John Benjamins.
- Purpura, J. (2004). *Assessing grammar*. Cambridge University Press.
- Ransom, L. (2009). Implementing the post-entry English language assessment policy at the University of Melbourne: Rationale, processes, and outcomes. *Journal of Academic Language and Learning*, 3(2), A13-A25. <https://journal.aall.org.au/index.php/jall/article/view/71>
- Read, J. (2008). Identifying academic language needs through diagnostic assessment. *Journal of English for Academic Purposes*, 7, 180-190. <https://doi.org/10.1016/j.jeap.2008.02.001>
- Read, J. (2015a). Issues in post-entry language assessment in English-medium universities. *Language Teaching*, 48(2), 217-234. <https://doi.org/10.1017/S0261444813000190>
- Read, J. (2015b). *Assessing English proficiency for university study*. Palgrave Macmillan.
- Read, J. (2016). Some key issues in post-admission language assessment. In J. Read (Ed.), *Post-admission language assessment of university students* (pp. 3-20). Springer International Publishing. https://doi.org/10.1007/978-3-319-39192-2_1
- Read, J., & Chapelle, C.A. (2001). A framework for second language vocabulary assessment. *Language Testing*, 18(1), 1-32. <https://doi.org/10.1177/026553220101800101>
- Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2017). How much vocabulary is needed to use English? Replication of Van Zeeland & Schmitt (2012), Nation, (2006), and

- Cobb (2007). *Language Teaching*, 50(2), 212–226.
<https://doi.org/10.1017/S0261444815000075>
- Scouller, K., Bonanno, H., Smith, L., & Krass, I. (2008). Student experience and tertiary expectations: Factors predicting academic literacy amongst first-year pharmacy students. *Studies in Higher Education*, 33, 167–178.
<https://doi.org/10.1080/03075070801916047>
- Shapiro, S., Farrelly, R., & Tomaš, Z. (2014). *Fostering international student success in higher education*. TESOL Press, NAFSA.
- Skinner, I., & Mort, P. (2009). Embedding academic literacy support within the electrical engineering curriculum: A case study. *IEEE Transactions on Education*, 52(4), 547–554. <http://dx.doi.org/10.1109/TE.2008.930795>
- The University of Auckland. (n.d.). *Diagnostic English Language Needs Assessment: Handbook for candidates at the University of Auckland*. Retrieved March 26, 2023, from <https://cdn.auckland.ac.nz/assets/delna/delna/documents/delna-handbook.pdf>
- The University of British Columbia. (n.d.). *English language competency*. Retrieved October 13, 2022, from Undergraduate Programs and Admissions: <https://you.ubc.ca/applying-ubc/requirements/english-language-competency/>
- Urmston, A., Raquel, M., & Tsang, C. (2012). Diagnostic testing of Hong Kong tertiary students' English language proficiency: The development and validation of DELTA. *Hong Kong Journal of Applied Linguistics*, 14(2), 60–82.
- Urmston, A., Raquel, M., & Aryadoust, V. (2016). Can diagnosing university students' English proficiency facilitate language development? In J. Read (Ed.), *Post-admission language assessment of university students* (pp. 87–112). Springer International Publishing. https://doi.org/10.1007/978-3-319-39192-2_5
- Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition*, 27(4), 567–595.
<https://doi.org/10.1017/S0272263105050254>
- Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science*, 9(2), 111–125.
<https://doi.org/10.1177/1745691613518075>
- Zumbo, B. (2007). Validity: Foundational issues and statistical methodology. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Psychometrics*, (vol. 26, pp. 45–79). Elsevier Science B.V. [https://doi.org/10.1016/S0169-7161\(06\)26003-6](https://doi.org/10.1016/S0169-7161(06)26003-6)
- Zumbo, B. (2021). *A novel multimethod approach to investigate whether tests delivered at a test centre are concordant with those delivered remotely online*. [Research Monograph]. Vancouver, BC: Paragon Testing Enterprises/UBC Paragon Research Initiative, University of British Columbia. <http://dx.doi.org/10.14288/1.0400581>