

## **Early English Language Assessment to Improve First-Year Student Success**

### **Dr. Penny Kinnear, University of Toronto**

Penny Kinnear currently works with the Engineering Communication Program at the University of Toronto where she focuses on the development and delivery of Professional Language support for a highly student body. She has a background in applied linguistics, second language and bilingual education and writing education. She is co-author of the book, "Sociocultural Theory in Second Language Education: An introduction through narratives." Her current research projects include a longitudinal study on professional identity development of Chemical Engineering students and a study of meaning-making language and behaviour in student design teams.

### **Dr. Micah Stickel, University of Toronto**

Dr. Micah Stickel (ECE) is Chair, First Year, in the Faculty of Applied Science & Engineering. He is also a Senior Lecturer in The Edward S. Rogers Sr. Department of Electrical & Computer Engineering. Dr. Stickel first came to the Faculty when he started as an undergraduate in 1993. Since that time, he has completed his BAsC (1997), MASc (1999), and a PhD (2006) - all with a focus on electromagnetics and the development of novel devices for high-frequency wireless systems. He has a great interest in engineering education and the use of technology to advance the student learning experience. He has been honoured with four departmental teaching awards and was selected as a New Faculty Fellow at the 2008 Frontiers in Education conference. In 2012, he was awarded the Early Career Teaching Award by the Faculty of Applied Science & Engineering.

### **Dr. Brian M. Frank, Queen's University**

Brian Frank is the DuPont Canada Chair in Engineering Education Research and Development, and the Director of Program Development in the Faculty of Engineering and Applied Science at Queen's University where he works on engineering curriculum development, program assessment, and developing educational technology. He is also an associate professor in Electrical and Computer Engineering.

### **Dr. James A. Kaupp, Queen's University**

Assessment and Quality Assurance Coordinator and Educational Researcher (Msc '06, PhD '12) at Queen's University, Kingston, Ontario, Canada in the Faculty of Engineering and Applied Science. Educational research interests include engineering education development, cultural change in higher education, higher-order thinking development and assessment, outcomes-based data-informed continuous improvement, authentic performance-based assessment, data analysis and information visualization

## Early English Language Assessment to Improve First-Year Student Success

### Abstract

This evidence-based practice paper reports on the first trial of two language screening and diagnostic instruments at two research intensive institutions, the University of Toronto in a large city and Queens University in a mid-sized city, partly in response to the growing linguistic diversity of the student body. The universities chose to use the Diagnostic English Language Needs Assessment (DELNA) jointly developed by the University of Auckland and the University of Melbourne, both of which had similar student demographics to the Canadian universities. DELNA has two parts, a screening and a diagnostic. The screening comprises a vocabulary section and a speed reading section that screen for academic vocabulary knowledge and academic literacy. The diagnostic is a written exercise that provides a finer grained analysis of student language use. The two universities chose to design and administer a diagnostic specific to an engineering faculty rather than the arts and humanities focus of the DELNA diagnostic. The screening revealed that although a majority of the students performed strongly on the vocabulary section of the screening, they performed much less strongly on the reading, indicating a weakness in academic literacy. The written diagnostic confirmed this and also revealed a number of different clusters of strengths and weaknesses, helping to dispel the myth, held by many students, that they just need help “fixing their grammar”. These preliminary findings indicate the need for further research, especially on the impact of identifying and providing relevant support for students on their ability to participate and perform in their courses.

### Problem

This project arose from the similar challenge faced by two research-intensive universities in Canada. Whereas monolingual English speakers had been the norm—students raised and educated in Canada in English—a new norm has emerged where a significant percentage of the students use a language other than English on a regular basis. These multilingual students, both Canadian citizens and international students, have complex linguistic histories. The students’ language learning experiences run along a continuum that ranges from studying English as a subject outside of Canada through completing formal elementary and secondary education in English in Canada in addition to formal and/or informal instruction in their ‘other’ language(s). Students with such histories are either required or choose to submit as a supplement to their high school English mark, English language proficiency test scores. The two most common measures are the International English Language Testing System (IELTS) or Test of English as a Foreign Language (TOEFL).

At the University of Toronto, a large research-intensive institution in a large city, many of the students enter with the required language proficiency scores. These students subsequently struggle in their classes, are perceived as drags on their first year teams and achieve lower marks and levels of understanding than they had been accustomed to in their high school classes. Queens University, a medium-sized research institution in a smaller city, has a much lower percentage of students who speak a language other than English at home—the vast majority were born in Canada and speak English as their primary or only language. However, the percentage of students for whom English is a second or third language is increasing. Design teams at Queens

University experienced similar problems when the majority of the students with strong English communication skills felt that their bilingual or multilingual teammates were not doing a fair portion of the work and the monolingual English speakers felt they had to edit and correct their teammates' writing.

Both universities are dealing with changing student demographics reflecting changes in Canada's population over the past several decades in addition to the recent trend in internationalization in Canadian higher education. The demographic change is not just from international students who make up an increasingly large proportion of the student bodies, but mirrors the ethnic, cultural and linguistic diversity and complex linguistic histories of the Canadian population. According to Statistics Canada's most recent available figures, immigrants make up 46% of the urban population where University of Toronto is located. Although the share of newcomers (recent immigrants and new Canadian citizens) settling in this urban area declined slightly since the last (2006) census, the area still received the largest share of newcomers, nearly one-third (over 380,000). Most of the immigrants between 2006 and 2011 have come from Asian and Middle Eastern countries with the second largest group coming from European countries.<sup>101</sup> As 61.2% of the immigrant population and 66.8% of the newcomer population speak one of Canada's official language and one or more non-official languages<sup>102</sup>, it would be surprising not to find this diversity reflected in the classroom. The ten most common non-official languages are in order of prevalence: Chinese (including Cantonese and Mandarin), Tagalog, Spanish, Punjabi, Arabic, Italian, German, Portuguese, Farsi and Polish.<sup>103</sup> The students accepted into the two programs are bright and hard-working but not all have had the opportunities to work within and with the English academic vocabulary and rhetorical discourses of the academy and, specifically, the engineering discipline.

### **Current practices**

The student and faculty frustrations have been defined in terms of English language proficiency, a concept that has been vaguely defined, primarily in terms of monolingual English written and oral performance standards. Historically, the question of proficiency has been addressed before admission through the use of standardized tests (IELTS, TOEFL, etc.) and high school English marks. If students met the standards set by the university, then they were deemed proficient. Post admission, it has been addressed through in-house English Proficiency Tests that involve students writing in response to a prompt. The student writing was judged holistically as proficient or not. If not, students would be placed in some kind of academic writing course. Neither the English language proficiency tests used as part of the application process nor the post-admission writing tests address the central challenge of learning to use the professional engineering discourse within an academic context. They also do not provide diagnostic information to help students improve their English as their primary learning tool. The IELTS and TOEFL scores that are used as admission criteria claim to provide a basis for predicting a student's readiness to handle the language of an academic course of study and their "*scores* are said to extrapolate to performance in real-life academic settings."<sup>104</sup> Two issues have been identified with the use of such tests. The first is the extent to which the extrapolation holds true "for the *actual language use*"<sup>105</sup> as the evidence has not been extensively investigated. Secondly, the language proficiency measures are not designed to provide diagnostic information about a student's language use post-admission. While both institutions use IELTS or TOEFL as part of admissions requirements, University of Toronto no longer uses an in-house designed writing

task. Queens University has continued with an in-house English Proficiency Test that serves as a graduation requirement, but does not explicitly provide diagnostic information.

At University of Toronto, although students are admitted into the program without caveats, every year a number of students struggle; some fail and leave the program (<5%), others become dependent on more linguistically competent peers or find ways to avoid language demanding tasks. The students are frustrated and the faculty is frustrated. At University of Toronto a program of support mechanisms has been cobbled together but the supports and the students do not connect regularly or systematically. Identification of students who would benefit from support has relied primarily upon the recommendation of an observant TA or writing tutor or the student seeking help after failing. Fear of stigmatizing or being stigmatized as “ESL” has kept faculty from recommending students to seek support and students from seeking support themselves. The issue has been reframed over the past five years as one of developing a student’s Professional Language in order to participate in learning and practicing engineering. This has helped to begin to change the perception of language support. However, the efficacy of these efforts has been hampered by the lack of an efficient, reliable and timely identification and diagnostic strategy or instrument as well as the voluntary nature and additional workload that participation in the support activities demands.

Queens University faces some similar issues; students who failed the English Proficiency Test were encouraged to enter an English support program, although its voluntary nature and additional workload limited participation. Since students are only required to pass the EPT by the time of graduation many who fail still continue working in team design courses, causing them to struggle with writing assignments and leading to friction on design teams.

### **Post enrollment screening for identification**

Both universities needed information that would allow them to address the unreliability of IELTS and TOEFL to extrapolate from scores to predict competence in actual language use contexts; in other words, to answer the question of why students who have high TOEFL or IELTS scores are inarticulate or unable to participate in academic activities. The schools also needed information that could be used to design the support necessary to enhance the development of academic and professional language. Fox’s findings from a study examining the use of diagnostic assessment concluded that such instruments can lead to more targeted instruction to make support of language development more efficient and effective<sup>[3]</sup>, something that would potentially be more useful than offering additional grammar instruction or 5-paragraph essay instruction. The schools needed to first identify students who might encounter difficulty in participating in their academic community. Second, the schools needed a way to better assess the needs of those students. Our universities were not the first to try and deal with this challenge. The University of Auckland and the University of Melbourne, both schools with similar demographics, confronted this task more than ten years ago and collaboratively developed the Diagnostic English Language Needs Assessment (DELNA)<sup>[4]</sup> (for further information about the DELNA see <http://www.delna.auckland.ac.nz/en.html>). The post enrollment language assessment would add a layer of screening similar to current math proficiency tests in common use among engineering programs. A critical difference between math and language is that math skills are more discreet and easier to isolate, identify and then instruct towards improvement, than learning to use a language to learn new content.

Read addresses what kinds of information a post-enrollment test may provide in his discussion of the development of the Diagnostic English Language Needs Assessment (DELNA) <sup>(9)</sup>. The DELNA is based on three constructs: academic language competence, academic language proficiency and academic literacies <sup>(9)</sup>. The screening phase of the DELNA comprises two sections, the first is a vocabulary task. Students must choose the correct definition of isolated academic vocabulary. The second is a timed reading task or cloze-elide procedure <sup>(9)</sup>. According to Read, these two tasks measure students' knowledge of academic language and the speed they can access that knowledge. The assumption is that if students score high on both of these measures they have a high degree of competence in academic language and solid academic literacies. These students would most likely not need support to succeed in an academic program. Students who do not score high on the screening measures would most likely need some kind of support. The Diagnostic phase of the DELNA comprising three tasks: listening, reading and writing, can provide more specific information that could guide support design. The DELNA Diagnostic, however, is targeted at a general university population and, as Read points out, focuses on academic language proficiencies and not academic literacies <sup>(9)</sup>. Given that both University of Toronto and Queens University are addressing the needs of a professional program with specific discourse norms and practices associated with the discipline such a generic diagnostic would not provide the information needed to enhance the development of academic literacies. As a result, the two universities agreed to collaborate on the design of a written diagnostic that could be used as the follow-up to the DELNA screening. This diagnostic would be administered to those students identified by the DELNA screening who did not exhibit solid academic language competence.

### **Development of an Engineering specific diagnostic**

Through a series of discussions, the several areas deemed critical to being able to participate in first-year engineering discourse were identified. These were: literal comprehension and production of information from a simple graph; interpretive and inferential production from the same graph in different scenarios, supported argumentation, and speed. In order to accomplish these tasks students would need to use vocabulary, syntax and rhetorical moves that became increasingly complex. By dividing the task into discrete levels that could be evaluated separately it was assumed that analytical scoring of student performance would provide evidence of student strengths and weaknesses that could be used to inform instruction. For example, did a student produce sentences free of syntax errors but without any logical organization or produce a series of logically connected but syntactically inaccurate sentences? The different patterns would demand different instructional materials and strategies.

Other criteria used in the design of the writing task included efficiency of administration and analysis. It was agreed that the entire task should take no more than 30 minutes. It was also agreed that it would be a paper-based task in order to provide the opportunity for students to use graphics in their responses, if they wished. This was important as graphics are an important part of engineering discourse practices. The diagnostic task was divided into three sections. The first required a literal reporting of the information found in a simple graph. As such it only required simple declarative sentences and vocabulary available in the graph itself. The second section asked for an interpretation of the information from the graph, a statement of a relationship of two sets of information found in the graph and a simple recommendation to an audience of their

peers. This section provided the opportunity for the use of more complex sentences (a comparison) and simple rhetorical argument structure, e.g. claim and evidence. The final section asked students to make inferences from the information in the graph and the information included in a short scenario and to communicate the outcome of that inference as a recommendation to be used by a person without a technical background. This section required the use of both complex syntax and rhetorical organization of paragraphs that made a supported argument, as well as accurate comprehension of the content (vocabulary) contained in the scenario and a lexical range that included formal but non-technical terms.

Once the content of the diagnostic had been confirmed, an analytic rubric was designed to be used by the assessors of the screening. It includes dimensions of argumentation, audience selection, inferencing, lexicogrammatical, word choice, concision, and content, plus one for speed. The detailed rubric is shown in Appendix A.

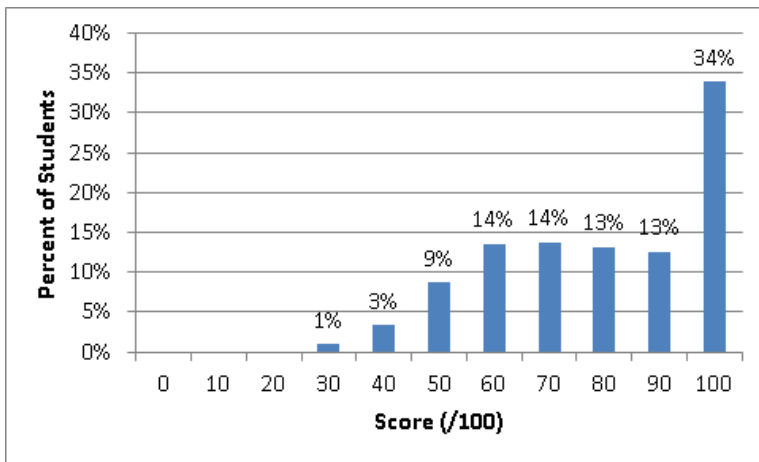
### **DELNA screening implementation and results**

At University of Toronto, 1224 first-year students wrote the DELNA screening test during the first two weeks of classes. Screening results were sent via email to each student together with an interpretation of their results and suggestions of resources and strategies students could access. Students who received a total DELNA score below the 60% cutpoint (a designation of Band 1) were asked to write the diagnostic part of the test during the first two weeks of October. At Queens University 761 students wrote the DELNA screening test during the first week of classes, and 96 students were flagged to write the diagnostic two weeks later based on receiving a score below the 60% cutpoint. This data is summarized in Table 1.

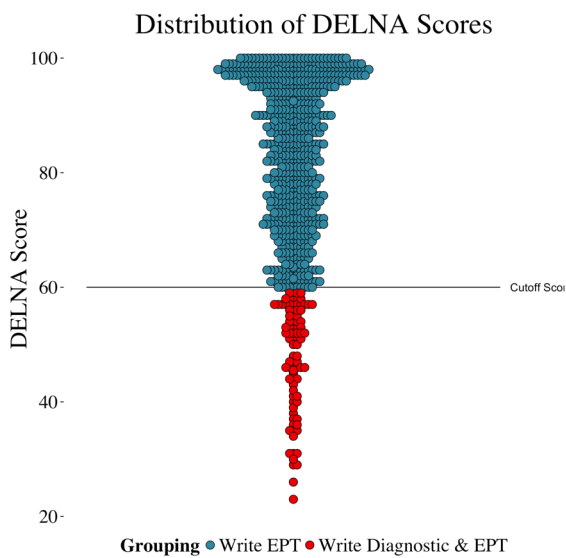
The DELNA screening score distributions for University of Toronto and Queens University are shown below in Figures 1 and 2, respectively.

**Table 1:** Summary of DELNA Screening Completion

<b>University</b>	<b># DELNA test written</b>	<b># scores &lt;60%</b>	<b>% of DELNA scored &lt;60%</b>	<b># diagnostics written</b>
University of Toronto	1224	366	30%	315
Queens University	761	96	13%	37



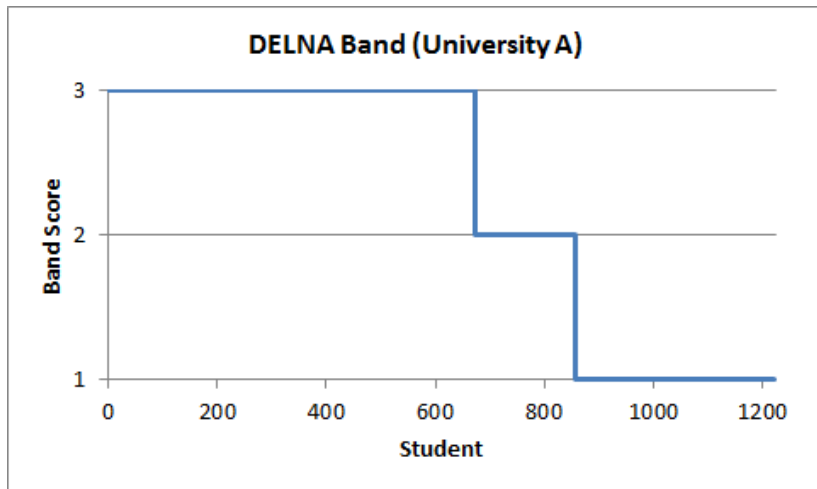
**Figure 1:** Distribution of DELNA screening diagnostic scores at University of Toronto.



**Figure 2:** Distribution of DELNA screening diagnostic scores at Queens University. EPT is the English Proficiency Test, which is a separate test required for all students at Queens University.

Figures 3a and b show the distribution of scores on the three Bands. Students in Band 1 with scores <60%, were those flagged to write the diagnostic. Students with scores  $\leq 60\%$  and <75%

make up Band 2, while students who scored  $\leq 74\%$  make up Band 3. In both cases Band 2 made up the smallest group.



Comment [PK1]: I don't have the original and I don't know how to take (University A) out and replace it with University of Toronto. Same with all of the figure below.

Figure 3a: Distribution of band scores on the DELNA screening diagnostic for University of Toronto

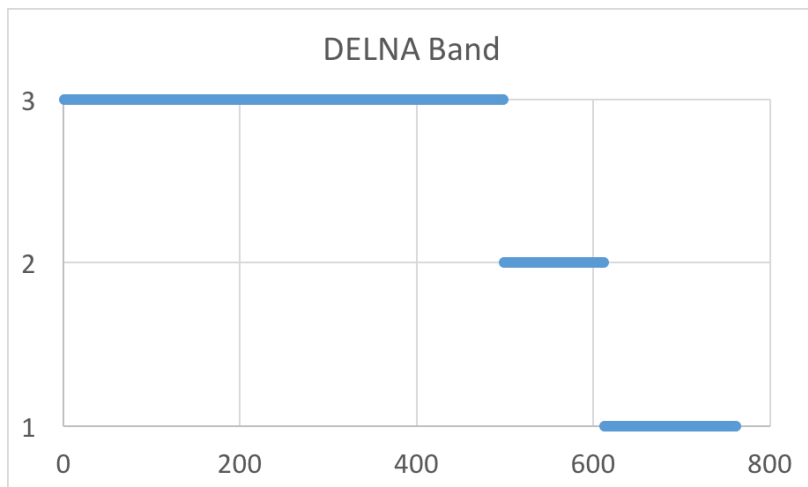


Figure 3b: Distribution of band scores on the DELNA screening diagnostic for Queens University

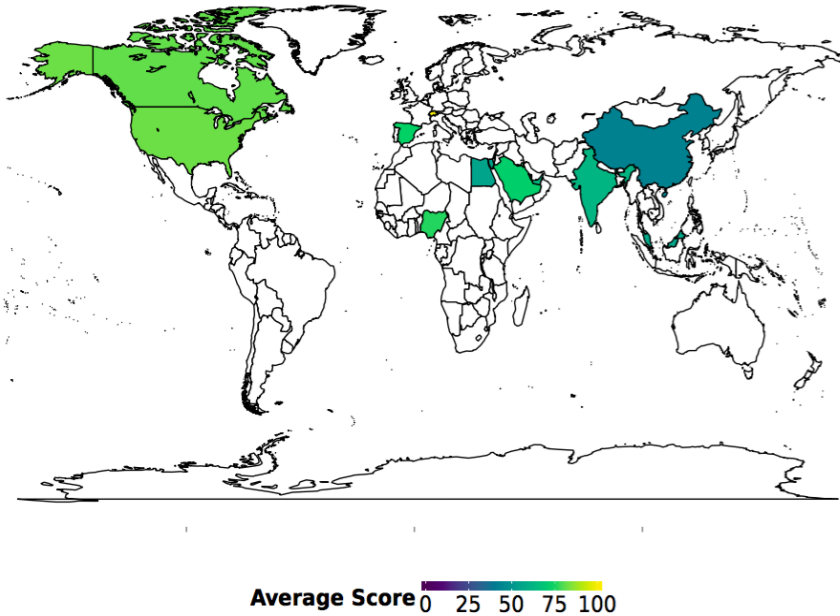


Table 2 summarizes how students at University of Toronto performed on the DELNA assessment based on the students' country at time of application. Figure 4 presents similar data for Queens University. Numbers drawn from countries in Africa and Europe are quite low, but there are a significant number from China, which has the lowest mean scores. It is likely that the variation in performance by home country reflects different educational histories. Often Chinese students study English only as a foreign language in English language classes. The exams are designed to measure their knowledge of the language, e.g. vocabulary, syntax rules, and ability to recognize grammaticality. Some students from other countries also studied English in similar contexts, however, many of the students from India and the Middle East attend English-medium schools or international schools where the language of instruction was English or have participated in International Baccalaureate programs. However, the fact that nearly one-quarter of students from Canada at University of Toronto perform at a Band 1 level indicates that it is important to provide additional support to both international and domestic students.

**Table 2:** Mean scores by University of Toronto students' country at time of application

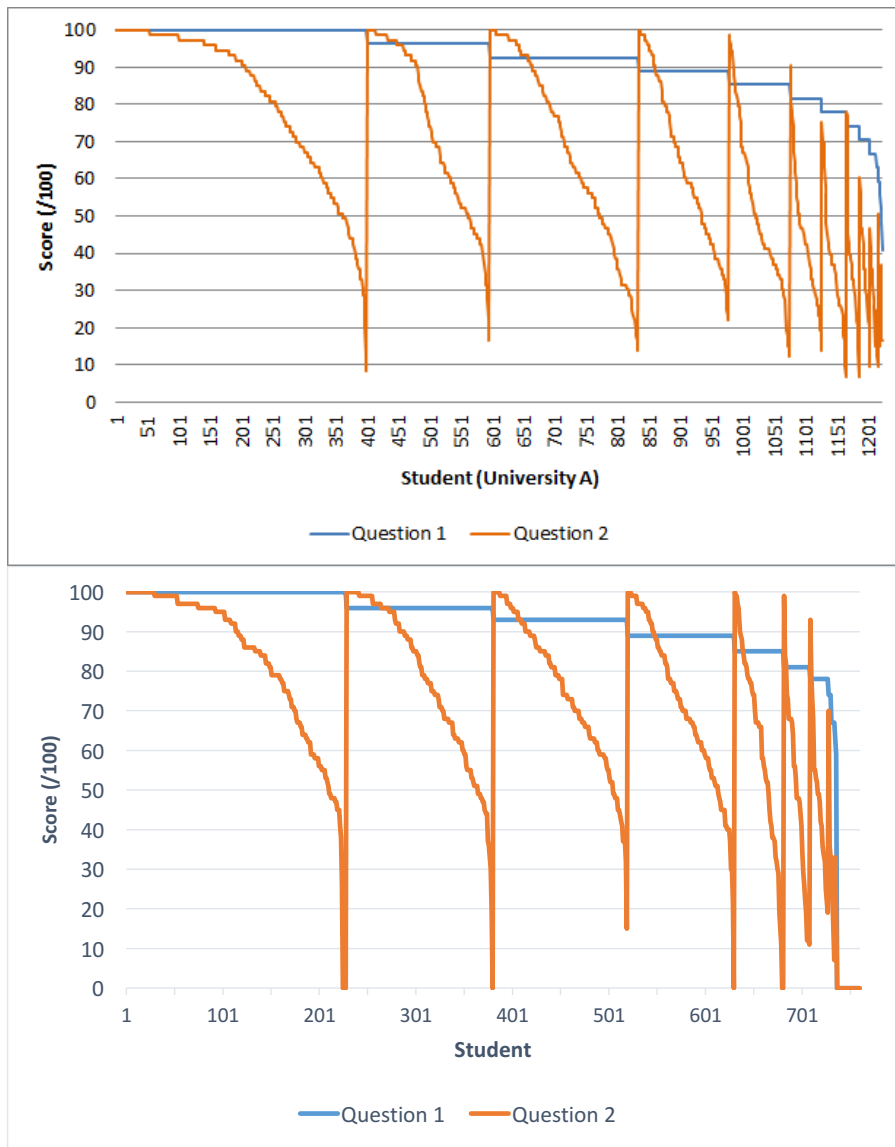
<b>University of Toronto</b>	<b>Count</b>	<b>Mean</b>	<b>Band 1 (&lt;60%)</b>	<b>Band 2 (&gt;60%, &lt;75%)</b>	<b>Band 3 (&gt;74%)</b>
<b>Canada</b>	939	78%	23.7%	14.8%	61.4%
<b>China</b>	128	55%	75.0%	14.8%	10.2%
<b>U.S.A.</b>	14	87%	7.1%	0.0%	92.9%
<b>United Arab Emirates</b>	14	83%	14.3%	21.4%	64.3%
<b>Pakistan</b>	8	77%	12.5%	12.5%	75.0%
<b>Hong Kong</b>	7	89%	14.3%	0.0%	85.7%
<b>Bangladesh</b>	6	72%	16.7%	50.0%	33.3%
<b>India</b>	6	80%	0.0%	50.0%	50.0%

## DELNA Distribution



**Figure 4:** Mean scores by Queens University students' country at time of application

Figure 5 shows individual student performance on the vocabulary (Question 1) and timed reading (Question 2), from high to low by vocabulary. It shows that there are a large number of students who did well on the vocabulary question, which required students to choose the correct definition of isolated academic vocabulary and had a mean score of 93% (University of Toronto) and 94% (Queens University), but quite poorly on question 2, a cloze-elide procedure requiring students to identify extraneous words that do not belong in a passage of text that had a mean score of 68% (University of Toronto) and 74% (Queens University). This seems to indicate either that the students were slow at accessing their academic vocabulary in order to make judgments about the meaning or lack of meaning of that vocabulary in an academic passage or possibly did not know the discourse patterns used in academic writing and thus were unable to make accurate judgments, or perhaps some combination of these. Further information would be necessary to make any accurate interpretation of these patterns.



**Figure 5:** Scores by student for vocabulary (Q1) and timed reading (Q2) on the DELNA screening diagnostic at University of Toronto (top) and Queen's University (bottom).

## Diagnostic implementation and results

At University of Toronto scoring the diagnostics was done by a team of five assessors knowledgeable of second language learning and writing. Consistency was established through a 2-hour benchmarking session. At Queens University scoring was done by the coordinator of the English Support for Engineers program and writing tutors, preceded by a 1-hour benchmarking session. In all, 315 diagnostics were analyzed at University of Toronto, and 37 at Queens University.

At Queens University about half of the 37 students writing the diagnostic received a score of 70 or higher and appeared to have no difficulty with the time limit. All students had sufficient time to at least start all the questions, and only 2/37 did significantly poorer on question 3 than the other two questions for reasons that appeared to be due to time. The mean score on each of the three questions was similar.

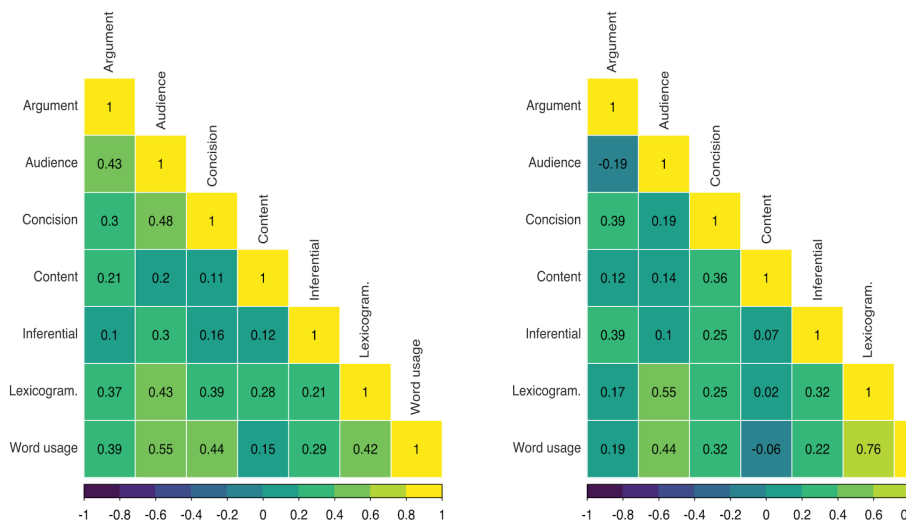
At University of Toronto a preliminary analysis of the diagnostic results identified four classes of support needs. Overall, timing did not appear to be a concern for most students with 85% of the group completing or attempting all three tasks, with only 1% not completing tasks 1 and 2. These classes of needs are as follows:

1. Students in the first group completed all three tasks with minimal errors that did not impede meaning and without issues in their inferencing and argumentation (rhetorical organization). These students had the necessary language knowledge but needed to develop their editing and proofreading skills and strategies.
2. Students in the second group were able to complete the first two tasks with minimal error and/or muddling of meaning or issues in their rhetorical organization, but were unable to complete the third task. These students appeared to need strategies to help them develop both reading and writing speed without sacrificing meaning.
3. Students in the third group completed the tasks but there were problems with understanding the meaning of what they had written because of multiple errors and/or rhetorical missteps. These students seemed to need both instruction and practice with syntax, rhetorical organization and vocabulary.
4. A fourth group had multiple errors that often muddled the meaning, issues with rhetorical organization and did not complete the three tasks. These students appeared to be the most in danger of being unable to successfully contribute to their team assignments or individual assignments or exams.

Hierarchical clustering using average distance was used at Queens University to identify common support needs. Three clusters of students were identified; students in cluster 1 did well on most of the test, to the point that no additional support would be recommended. Students in cluster 2 generally had scores of 1 or 2 on most dimensions, and need would benefit from support in multiple areas. Students in cluster 3 did relatively well on audience and lexicogrammar, but had issues with argumentation, and inferencing. These two dimensions are among the most complex of the expectations on the diagnostic.

Figure 6 shows the correlations between each of the dimensions at each University. Lexicogrammatical, audience, and word usage are the most strongly correlated with each other.

Argumentation, inferencing, and concision moderately correlate with each other, particularly at Queens University. Since a characteristic of students in cluster 3 at Queens University above was low scores in argumentation and inferencing, the correlation scores appear to support a link between these expectations.



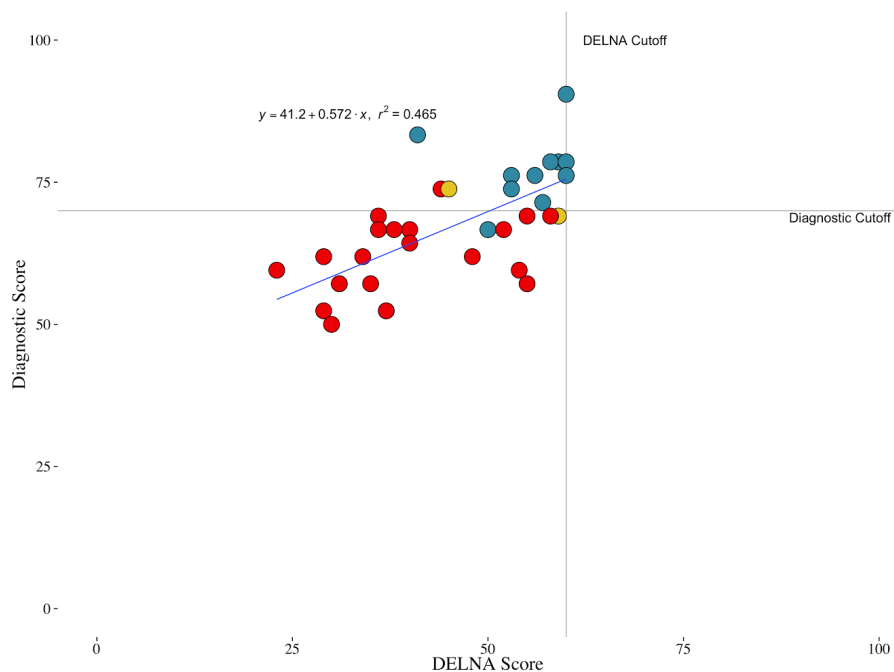
**Figure 6:** Correlations between dimensions for University of Toronto (left) and Queens University (right).

A simple correlational analysis was done for the University of Toronto cohort to assess any possible relationship between the DELNA screening diagnostic and the first-year engineering design and technical courses. The summary of this analysis is presented in Table 3. As might be expected reasonable correlations were found between both aspects of the screening diagnostic and the course which focuses on engineering design and communications. No significant correlations were identified between the screening diagnostic and the technical courses, other than the weaker relationships found with the computer programming course.

The relationship between total scores on the DELNA and diagnostic for the Queens University cohort are shown in Figure 7 below, with an  $R^2$  of 0.47. The color of the dots represents students' scores on the required English Proficiency Test (EPT). Though that test was not the focus of this work, it is useful to note that, of the students flagged on the DELNA, most students who received a failing score on the EPT (by a different grader) were also below the cutoff on the diagnostic.

**Table 3:** Correlations between DELNA screening diagnostic and first-year course performance.

	DELNA Vocab.	DELNA Timed Reading	Mechanics	Eng. Design I	Calculus I	Linear Algebra	Comp. Prog.
DELNA Vocab.	1.00	0.54	0.12	0.43	0.08	0.08	0.23
DELNA Time Reading		1.00	0.05	0.48	-0.03	-0.02	0.21



**Figure 7:** Relationship between Queens University scores on the DELNA, diagnostic, and locally required English Proficiency Test.

## Implications for action

It could be argued that one of the most important results of the granularity that our diagnostic has provided is the ability to counter the claim of many of our students (and some of our faculty) that students simply need “more English” or “more grammar instruction”. Our diagnostic shows that neither of those solutions would help the students more successfully participate in the classroom activities. Instead we can begin to design more specific interventions and support activities, ideally, in collaboration with course instructors. The screening and the diagnostic were piloted for the first time in the 2015-2016 academic year in order to get more reliable information first, about the number of students who would potentially struggle. Second, the programs wished to obtain more detailed information about the nature of their struggles with academic and engineering English in order to inform possible curricular changes and other support mechanisms.

Some of these initiatives at University of Toronto included the Professional Language Development coordinator making herself available to advise students based on the results of the diagnostic. First-year design course workshops were announced only to those students identified in Band 1. Workshop foci were adjusted to make a better match between identified needs and content. For example, “Mid-term Preparation” was changed from strategies for reading and interpreting multiple choice questions to strategies for reading the case study, used as the basis of the multiple choice questions, through the lens of engineering concepts. This addressed some of the academic literacy skills identified in the Screening and the Diagnostic. Communication Instructors who work with the two first-year courses were made aware of the students who scored in Band 1 and given access to their diagnostic results in order to better inform their interactions with the students on their written assignments. These are only preliminary responses to the screening and diagnostic process. Currently, plans are being made to address the challenges of earlier and more efficient communication of the diagnostic results to both students and instructors. Plans are also being made to more efficiently collaborate with students through strategy sessions and their course assignments.

At Queens University the 24 students whose diagnostic scores were <70%, or who received a speed score <3, were asked to attend a support session that provided assistance and feedback on writing activities for the design course. The approach was to provide support for academic activities rather than assign remedial work. Students were also provided assistance to help them complete the English Proficiency Test (required for graduation). Students who did not pass the EPT offered at the beginning of the semester were also invited to these support sessions.

However, participation in these activities at both universities is not mandatory. This means that some students will never take advantage of the supports offered. However, as part of a long-term strategy that has attitudinal change as one of its goals, keeping participation voluntary allows the emergence of new attitudes based on the satisfaction and demonstrated success of students who participate. Students telling their friends that visiting a tutor, participating in a library research workshop or a concision and conjunctions workshop helped them do better in their classes supports our efforts more effectively than adding yet another required activity to the already packed schedules of our students. Fox reported a shift in student attitudes toward tutoring and support for language development within her university’s engineering program from a few hours

a week availability to a 24/7 hub of teaching and learning, a cultural shift, none of which was ever made mandatory. <sup>[5]</sup>

### What have we learned

We have a significant number of students, both domestic and international, who appear to struggle with the transition from studying language as a subject to using language to study engineering and participate in the engineering discourse. This is reflective of the internationalization of higher education and recruitment efforts as well as the increasing diversity of our domestic population. We can see that language proficiency tests only provide one kind of measure but are not particularly helpful in identifying student needs when transitioning from language as subject to language as medium of instruction or language as practice. The DELNA screening appears to reliably identify students who would benefit from a finer grained analysis of their academic language and literacy strengths and weaknesses. The diagnostic appears to identify clusters of students with particular needs, including some who need support with lexicogrammatical issues and word choice while others need support with the more complex elements of argumentation, concision, and inferencing. Further analysis of our results will help us to revise and refine both the diagnostic and the analytic rubric. Does this have significance for policies around the use of language proficiency standards? Admission screening practices? These questions can only be answered with further research and analysis.

Linguistic, cultural and educational preparation diversity are a reality. This diversity has an impact on student learning in our programs. We believe we have made some progress with a more reliable identification practice. This has allowed us to counter some student (and faculty) perceptions of students as second language learners who just need more grammar instruction with a more nuanced description of what students need to focus on and strategies for using their language knowledge. Although we have improved the support for identified students we have not been as timely or complete as we would like or need to be. We need to develop more efficient methods of gathering and distributing specific information gleaned from the diagnostics directly to the students to help them become aware of what they need to do and resources they can use to address their needs. Given the number of students we have been able to identify, we also need to consider how we can develop our own capacity to support these students in order for them to take full advantage of the teaching and learning contexts they are participating in.

### References

- [1] Statistics Canada, Immigration and Ethnocultural Diversity, Catalogue Number, 99-010-X, May 2013, available: <https://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-010-x/99-010-x2011001-eng.cfm>
- [2] L. Brooks and M. Swain, "Contextualizing performances: Comparing performances during TOEFL iBT™ and real-life academic speaking activities," *Language Assessment Quarterly*, vol. 11, pp. 353-373, 2014.
- [3] J. D. Fox, "Moderating top-down policy impact and supporting EAP curricular renewal: Exploring the potential of diagnostic assessment," *Journal of English for Academic Purposes*, vol. 8, pp. 26-42, 2009.
- [4] J. Read and J. v. Randow, "A university post-entry English language assessment: Charting the changes," *International Journal of English Studies*, vol. 13, pp. 89-110, 2013.
- [5] J. Read, "Issues in post-entry language assessment in English-medium universities," *Language Teaching*, vol. 48, pp. 217-234, 2014.
- [6] J. Fox, private communication, January 2015.



## Appendix A: Diagnostic Rubric

<b>Question 1 - Graphical literal interpretation</b>				
Criteria		<b>Not demonstrated (1)</b>	<b>Developing (2)</b>	<b>Proficient (3)</b>
Accuracy	Content	Misinterprets/misunderstands meaning or not attempted or Includes irrelevant information not included in the prompt	Incomplete description of data, may include some irrelevant data	Complete description of data without misinterpretation or misunderstanding
	Lexico-grammatical	Error rate and type make writing incomprehensible; spelling, capitalization, punctuation; word form; subject-verb agreement; verb tense choices; singular/plural; syntax and clause ordering	Error rate and type force reader to read the selection multiple times to grasp meaning; leaves a reader unsure of intended meaning	Meaning is clear although there may be a few minor errors, e.g. spelling, capitalization, punctuation; subject-verb agreement; verb tense choices
<b>Question 2- Graphical inferential response to an audience</b>				
Criteria		<b>Not demonstrated (1)</b>	<b>Developing (2)</b>	<b>Proficient (3)</b>
Accuracy	Lexico-grammatical	Error rate and type make writing incomprehensible; spelling, capitalization, punctuation; word form; subject-verb agreement; verb tense choices; singular/plural; syntax and clause ordering	Error rate and type force reader to read the selection multiple times to grasp meaning; leaves a reader unsure of intended meaning	Meaning is clear although there may be a few minor errors, e.g. spelling, capitalization, punctuation; subject-verb agreement; verb tense choices
Rhetorical	Inferential/ Interpretation	Repeats content from prompt w/o establishing an interpretive relationship e.g. so, because, therefore, if-then	Includes relevant information and interpretation but not presented in any recognizable order or pattern w/o specific reason or justification; lacks because, therefore, if-then, etc. clauses	Provides an interpretation with relevant information in a known to new, part to whole, sequential or relational order and at least one reason or justification from the content
Vocabulary	Word usage	Multiple errors in word use and formation distorting meaning	Overuse of prompt vocabulary, inappropriate use of synonyms,	Appropriate vocabulary choices with minor word form errors

			multiple word form errors, but comprehensible with effort	
	Audience appropriateness	No evidence of deliberate word choices, e.g. mixed register and levels of technicality; inaccurate choice of 'everyday' synonyms	Overuse of technical vocabulary, inconsistent attempt to use 'everyday' language	Some technical vocabulary included but with appropriate explanation in 'everyday' language or only 'everyday' language that substitutes accurate synonyms for technical terms
Concision		Circumlocution, repetitive, irrelevant information, overuse of prepositional phrases in place of adjectives	Overuse/inappropriate use of complex sentence structures, some irrelevant information, indiscriminate use of qualifiers e.g. very, excellent, extremely	Simple, compound and complex sentences used in balance (to best serve meaning?), relevant information, qualifiers used to add precision
<b>Question 3 - Reading inferential argument to an audience</b>				
Criteria		<b>Not demonstrated (1)</b>	<b>Developing (2)</b>	<b>Proficient (3)</b>
Accuracy	Content	Misinterprets/misunderstands meaning or not attempted	Partial description of data	Most data described
	Lexico-grammatical	Error rate and type make writing incomprehensible; spelling, capitalization, punctuation; word form; subject-verb agreement; verb tense choices; singular/plural; syntax and clause ordering	Error rate and type force reader to read the selection multiple times to grasp meaning; leaves a reader unsure of intended meaning	Meaning is clear although there may be a few minor errors, e.g. spelling, capitalization, punctuation; subject-verb agreement; verb tense choices
Rhetorical	Argument	States a claim w/o evidence or justification	States a claim, provides at least one justification but does not explicitly or implicitly link with claim; connectives do not reflect the relationship between ideas	States a claim, uses evidence in either an inductive or deductive organization with connectives that describe the relationship between the ideas
Vocabulary	Word usage	Multiple errors in word use and formation distorting meaning	Overuse of prompt vocabulary, inappropriate use of synonyms, multiple word form errors, but comprehensible with effort	Appropriate vocabulary choices with minor word form errors
	Audience appropriateness	No evidence of deliberate word choices, e.g. mixed register and levels of technicality; inaccurate choice of 'everyday' synonyms	Overuse of technical vocabulary, inconsistent attempt to use 'everyday' language	Some technical vocabulary included but with appropriate explanation in 'everyday'

				language' or only 'everyday' language that substitutes accurate synonyms for technical terms
Concision		Circumlocution, repetitive, irrelevant information, overuse of prepositional phrases in place of adjectives	Overuse/inappropriate use of complex sentence structures, some irrelevant information, indiscriminate use of qualifiers e.g. very, excellent, extremely	Simple, compound and complex sentences used in balance (to best serve meaning?), relevant information, qualifiers used to add precision
<b>Overall</b>				
		<b>Not demonstrated (1)</b>	<b>Developing (2)</b>	<b>Proficient (3)</b>
Speed		Completes first two questions but does not address third question in time	Completes first two questions and most of third in time	Completes all required questions in allocated time