



Higher Education
Quality Council
of Ontario

An agency of the Government of Ontario

Measuring Critical-thinking Skills of Postsecondary Students

Ross Finnie, Michael Dubois, Dejan Pavlic,
Eda Suleymanoglu (Bozkurt)
Education Policy Research Initiative



Published by

The Higher Education Quality Council of Ontario

1 Yonge Street, Suite 2402
Toronto, ON Canada, M5E 1E5

Phone: (416) 212-3893
Fax: (416) 212-3899
Web: www.heqco.ca
E-mail: info@heqco.ca

Cite this publication in the following format:

Finnie, R., Dubois, M., Pavlic, D. & Suleymanoglu (Bozkurt), E. (2018). *Measuring Critical-thinking Skills of Postsecondary Students*. Toronto: Higher Education Quality Council of Ontario.



The opinions expressed in this research document are those of the authors and do not necessarily represent the views or official policies of the Higher Education Quality Council of Ontario or other agencies or organizations that may have provided support, financial or otherwise, for this project. © Queens Printer for Ontario, 2018

Executive Summary

The skills — or competencies — that individuals develop play a fundamental role in determining their labour market opportunities and more generally their life chances. At a broader level, the development of skills is critical to economic prosperity, social progress and overall quality of life in Canada.

While skills can be developed throughout an individual's life, the role of postsecondary education (PSE) in the development of these skills is of significant interest to skills stakeholders such as government agencies, corporations and PSE institutions. The relevant questions are:

- Which skills are of value in the labour market and to other life outcomes?
- Which of these skills can be developed or improved in PSE and how?

In this context, the Postsecondary and Workplace Skills (PAWS) project was created as a collaborative partnership involving a large Ontario college, a large Ontario university, the Education Policy Research Initiative (EPRI) and the Higher Education Quality Council of Ontario (HEQCO), with additional funding and support from Employment and Social Development Canada (ESDC).

PAWS focuses on one particular set of skills, namely those related to critical thinking. The goals of the project are to:

1. Measure and compare entering and graduating students' critical-thinking levels.
2. Identify the relationships between critical-thinking scores and a range of student and program characteristics.
3. Identify the value of critical-thinking skills in the labour market.

This report addresses the first two questions. The second phase of the project will involve relating critical-thinking skill levels to labour market outcomes by linking the data gathered for the project to tax data held by Statistics Canada. The results of the second analysis will be presented in a subsequent report.

To measure critical thinking, PAWS uses the HEIghten Critical Thinking assessment — a 45-minute online test developed by the Educational Testing Service (ETS) to measure the *analytic* and *synthetic* skills of college and university students. During the 2016–2017 academic year, the HEIghten assessment was administered to a selection of students in the entering and graduating classes at each institution.

Using descriptive and regression approaches, the main results of the analysis are as follows:

College results:

- There is no statistically significant difference in the HEIghten scores between the graduating and entering class.

- Students who are male, whose first language is English or French, who are not visible minorities, who did well on entry assessment tests (reading, writing and math) and who spent more time on the test tend to record higher critical-thinking scores in both classes.
- Students who were age 24–26 at entry tend to score higher than those who were age 18 and under at entry in the entering class. Students aged 22–23 at entry tend to score higher than those aged 18 and under at entry in the graduating class.
- Scores appear to vary between programs. These differences are relatively small for the entering class, but are larger for the graduating class. Students in community and justice programs tend to score lower than students in other programs, the only exception being the scores of business students in the entering class.
- Based on a regression analysis, the patterns mentioned above still hold when differences in student and program characteristics, assessment scores and time spent on the test are taken into account, but the magnitudes of score differences by the various characteristics decrease, with the greatest effect coming from the assessment scores.
- Based on a regression analysis that combines the scores of entering and graduating students, differences in skill levels between the graduating and entering classes vary across age and programs, with students age 22–23 showing a larger difference in skills between the graduating and entering classes than other age groups, and community and justice studies showing a smaller difference than other programs. The differences in scores between classes are larger for those who scored within the second and sixth category of the reading assessment scores, smaller for those who scored in the second and third quartiles of the writing assessment score distribution, and larger for those who scored in the second and third quartiles of the math assessment score distribution compared to those who scored in the bottom quartile. No other differences between the graduating and entering classes across the range of student characteristics, program and the other variables included in the analysis are statistically significant.

University results:

- Overall, the critical-thinking scores of graduating students were 4.8 points higher than entering students, 168.3 and 163.5, respectively.
- For the entering class, students who score higher on average are male, have higher high school GPAs, communicate better in English, are not a visible minority, are studying engineering and put more effort into the test.
- For the graduating class, students who record higher scores have higher high school GPAs, are not a visible minority and put more effort into the test.
- The regression analysis shows that the score patterns generally do not change when all other factors (student characteristics, faculty and effort) are taken into account, except for gender in the

graduating class: taking into account other student characteristics, faculty and the effort put into the test, male students tend to score higher than their female counterparts.

- The gap in critical-thinking scores between the graduating and entering classes is smaller for Engineering students than Arts and Science students and larger for students who tried their best than for students who did not. No other differences between the classes across the student characteristics and the other variables are statistically significant.

We caution, however, that comparisons of the entering and graduating students included in this analysis, which represent two different sets of students assessed at a single point in time, may be driven by various selection issues discussed at length in this report.

Given these limitations, the report provides suggestions for an alternative approach. In particular, we suggest adopting a longitudinal approach where the same students would be followed from their entry into PSE to the completion of their studies, while assessing their skills at multiple points in time. This would provide measures of skill levels, changes in skill levels during PSE and information regarding how these changes may be related to a range of student characteristics and schooling experiences.

Nevertheless, the PAWS project begins to fill an important gap in our understanding of skills in the PSE context by measuring skill levels of entering and graduating students, while also developing an initial analytical framework that could be useful for future research of this type.

Table of Contents

Executive Summary.....	2
1. Introduction.....	8
2. Data	10
2.1 The HEIghten Critical Thinking Assessment	10
2.2 The College, University and Administrative Data Provided	11
2.3 Sample Selection	11
<i>College</i>	11
<i>University</i>	12
2.4 Variables Included in the Analysis.....	12
<i>College</i>	12
<i>University</i>	13
3. Methodology	14
3.1 Descriptive and Regression Analysis	14
3.2 Limitations	14
4. Findings — College	16
4.1 Summary Descriptive Analysis	16
4.2 Full Descriptive Analysis	19
<i>Test Scores by Class</i>	19
<i>Test Scores by Selected Characteristics</i>	20
4.3 Regression Analysis	26
<i>Entering Cohort</i>	26
<i>Graduating Cohort</i>	29
<i>Cross-cohort Comparison</i>	31
5. Findings – University	32
5.1 Summary Descriptive Analysis	32
5.2 Full Descriptive Analysis	35
<i>Test Scores by Class</i>	35
<i>Test Scores by Selected Characteristics</i>	36

5.3	Regression Analysis	43
	<i>Entering Cohort</i>	43
	<i>Graduating Cohort</i>	45
	<i>Cross-cohort Comparisons</i>	46
6.	Conclusion	47
	References	50

List of Figures

Figure 1: Kernel Density of HEIghten Scores by Cohort, College	16
Figure 2: Distribution of HEIghten Scores by Cohort, College	17
Figure 3: Violin Plots of HEIghten Scores by Cohort, College	20
Figure 4: Violin Plots of HEIghten Scores by Cohort and Sex, College	20
Figure 5: Violin Plots of HEIghten Scores by Cohort and Age, College	21
Figure 6: Violin Plots of HEIghten Scores by Cohort and Language, College	22
Figure 7: Violin Plots of HEIghten Scores by Cohort and First Generation Status, College	22
Figure 8: Violin Plots of HEIghten Scores by Cohort and Visible Minority Status, College	23
Figure 9: Violin Plots of HEIghten Scores by Cohort and Program, College.....	24
Figure 10: Violin Plots of HEIghten Scores by Cohort and Reading Assessment Score, College.....	24
Figure 11: Violin Plots of HEIghten Scores by Cohort and Writing Assessment Score, College	25
Figure 12: Violin Plots of HEIghten Scores by Cohort and Math Assessment Score, College.....	25
Figure 13: Scatter Plot of HEIghten Scores and Time Spent on the Test by Cohort, College	26
Figure 14: Kernel Density of HEIghten Scores by Cohort, University	32
Figure 15: Distribution of HEIghten Scores by Cohort, University.....	33
Figure 16: Violin Plots of HEIghten Scores by Cohort, University.....	35
Figure 17: Violin Plots of HEIghten Scores by Cohort and Sex, University	36
Figure 18: Violin Plots of HEIghten Scores by Cohort and High School GPA, University	38
Figure 19: Violin Plots of HEIghten Scores by Cohort and Language, University.....	39
Figure 20: Violin Plots of HEIghten Scores by Cohort and Visible Minority Status, University.....	39
Figure 21: Violin Plots of HEIghten Scores by Cohort and Faculty, University	40
Figure 22: Violin Plots of HEIghten Scores by Cohort and Test Completion, University	41
Figure 23: Violin Plots of HEIghten Scores by Cohort and Whether Students Tried Their Best, University	42
Figure 24: of HEIghten Scores by Cohort and Effort, University.....	43

List of Tables

Table 1: Descriptive Statistics, College.....	18
Table 2: Descriptive Statistics, University	34

1. Introduction

The skills — or competencies — that individuals develop play a fundamental role in determining their labour market opportunities and life chances more generally. At a broader level, the development of skills is critical to economic prosperity and social progress as well as the overall quality of life in Canada.

Recent thinking, however, motivated in substantial part by surveys of employers (Barrington & Casner-Lotto, 2006; Business Council of British Columbia, 2010; Business Council of Canada, 2016; Canadian Association of Career Educators and Employers, 2013; and Royal Bank of Canada, 2018) has prompted policy makers and various education stakeholders to consider a broader set of skills than in the past, when skills were typically discipline-specific knowledge and competencies associated with a given credential or occupation (see also Gomathy, 2017 and Johnson, 2009).

The new skills of interest include transferable skills. While essential skills (such as basic literacy and numeracy) have long been recognized as important, higher-order cognitive skills (such as critical thinking, problem solving and advanced communication) are now finding their way onto such lists. Lists of desired skills also include what have traditionally been referred to as soft skills, or even personal attributes. These skills are now known as non-cognitive skills by economists, social-emotional skills by psychologists, or 21st century skills by others to reflect the broad sets of skills that individuals need going forward.

While skills can be developed throughout an individual's life, the role of postsecondary education (PSE) in the development of these skills is of significant interest to skills stakeholders. The relevant questions are:

- Which skills are of value in the labour market and to other life outcomes?
- Which of these skills can be developed or improved in PSE and how?

In this context, the Postsecondary and Workplace Skills (PAWS) project was created as a collaborative partnership involving a large Ontario college, a large Ontario university, the Education Policy Research Initiative (EPRI) and the Higher Education Quality Council of Ontario (HEQCO), with additional funding and support from Employment and Social Development Canada (ESDC).

PAWS focuses on one particular set of skills, namely those related to critical thinking. The goals of the project are to:

1. Measure and compare entering and graduating students' critical-thinking levels.
2. Identify the relationships between critical-thinking scores and a range of student and program characteristics.
3. Identify the value of critical-thinking skills in the labour market.

This report addresses the first two questions. The second phase of the project will involve relating critical-thinking skill levels to labour market outcomes by linking the data gathered for the project to tax data held by Statistics Canada.

To measure critical thinking, PAWS uses the HEIghten Critical Thinking assessment — an online test that measures the analytic and synthetic skills of college and university students (see Appendix A). During the 2016–2017 academic year, the HEIghten assessment was administered to a selection of students in the entering and graduating classes at both participating institutions, which then transferred the dataset containing the assessment results along with student and program characteristics drawn from their administrative data to EPRI.

This report presents the results of an analysis of the relationship between various student/program characteristics and critical-thinking scores, as well as comparisons between the entering and graduating cohorts.

We find that college students in their final year of PSE record, on average, similar critical-thinking scores to those recorded by students in their first year of study, while university students from the graduating class record higher critical-thinking scores than students in the entering class.

Using descriptive and regression approaches, we find that students' critical-thinking scores do indeed vary by a range of student and program characteristics, and in a limited set of cases, differences in scores between the entering and graduating classes also vary.

However, we caution that differences between the entering and graduating students (which represent two different sets of students assessed at a single point in time) may be driven by various factors. For instance, critical-thinking scores may be influenced by differences in the characteristics of the students in the entering year versus the graduating year, by differences in the samples of students who took the test in the two years stemming from how students were recruited for the tests at each institution, by any underlying differences in the students in the two classes at each institution and potentially other factors as well.

We discuss these issues at length in this report and provide suggestions for an alternative approach. In particular, we suggest adopting a longitudinal approach where the same students would be followed from their entry into PSE through to the completion of their studies, and have their skills assessed at multiple points during this time. This would provide measures of skill levels, changes in skill levels during PSE, and information on how these changes may be related to a range of student characteristics and schooling experiences.

In the meantime, the PAWS project begins to fill an important gap in our understanding of skills in the PSE context by measuring skill levels of entering and graduating students, while also developing an initial analytical framework that could be useful for future research of this type.

The next section of the report provides a description of the data used in the analysis — including the HEIghten instrument used to measure critical-thinking skills — along with the other data provided both through the test and from participating institutions. This is followed by a section on methodology, which also outlines the limitations of the analysis as mentioned above.

The main results come next. This starts with a summary descriptive analysis based on the sample characteristics of the entering and graduating classes, including mean HEIghten scores for the two classes at each institution across the range of student characteristics and other variables included in the analysis, as well as a discussion of the different characteristics of the entering and graduating classes at each institution. This is followed by an extended descriptive analysis that looks at a broader range of summary statistics for the HEIghten scores, including mean and quartile scores for the two classes on a variable-by-variable basis. The descriptive results are first presented for college students and then university students. In the regression analysis, which follows, critical-thinking scores are related to the range of variables included in the analysis in different combinations. First, only student characteristics are included; then program/faculty characteristics are added; then — for the college sample — scores for reading, writing and math assessments administered at entry are included; and, finally, measures of the time and effort spent on the HEIghten test are added. These regression results are presented first for the entering and then for the graduating college cohorts, followed by the entering and graduating university cohorts. Regression results are also presented for both cohorts combined to test for whether the differences in critical-thinking scores between the two classes vary across student characteristics, program/faculty and effort put into the test.

Finally, the paper concludes with a short summary of the report, a concise list of the main findings and a recommendation for a longitudinal research framework for future work.

2. Data

2.1 The HEIghten Critical Thinking Assessment

The HEIghten Critical Thinking assessment is an online survey created by the Educational Testing Service (ETS), a non-profit organization of educators, researchers and assessment developers based in Princeton, New Jersey and committed to improving education by providing assessment tools.

The assessment was developed expressly for the purpose of measuring the critical-thinking skills of American college and university students. It was created following a comprehensive review of both the existing literature and 10 instruments previously used to measure critical thinking at the college level. Based on this review, HEIghten was designed to examine two equally weighted aspects of critical thinking: analytic skills and synthetic skills. The assessment is completed online and takes approximately 45 minutes.

Analytic skills are measured through questions that ask a student to analyze an argument's structure, identify assumptions or flaws in reasoning and evaluate the evidence used to support the argument presented in short texts they are given to read. Synthetic skills are measured by asking students to select information to support an argument again based on a short text they are given to read, and to identify the conclusions or implications that arise from the text (see sample questions in Appendix A.1)

Students receive scores for both components of the test, which are then combined to yield a total score from 150 to 180, higher scores corresponding to higher assessed levels of critical-thinking skills. Three critical-thinking proficiency levels are identified: developing (150–161 points), proficient (162–172 points) and advanced (173–180 points). Appendix A.2 describes the skills associated with each performance level.

In the spring of 2015, a team of researchers at ETS conducted pilot tests involving 3,036 students across 35 American PSE institutions comprised of both two- and four-year colleges and “demonstrated [that HEIghten Scores] were reasonably correlated with SAT scores, high school grade point average (GPA), and college GPA, and were able to detect cross-sectional performance difference between freshmen and seniors” (Liu, Mao, Frankel & Xu 2016).

The HEIghten assessment was selected by HEQCO in consultation with the participating college and university as the tool for measuring critical thinking in the project. The HEIghten assessment tool also collects other information on students, some of which was used in the analysis to complement the administrative data provided by the two PSE institutions.

2.2 The College, University and Administrative Data Provided

The two PSE institutions included in the PAWS project were recruited by HEQCO based on a number of factors including: their extensive experience with student testing; their general capacity to administer the HEIghten test and to provide additional administrative data; and finally their willingness to participate in the project.

It should not be assumed that the college or the university are in any way representative of the PSE sector as a whole, or of the college and university student populations in the province or country more generally.

As part of their commitment to the PAWS project, these institutions provided data on their students in addition to the HEIghten data. More specifically, the college provided a student roster file with key student characteristics, a term file that could be used to obtain student program information, plus assessments of their reading, writing and math ability carried out at entry to the college. Variables provided by the university used in the analysis include sex and faculty.

2.3 Sample Selection

At each institution, the HEIghten test was administered to two groups of students: one from the entering class and another from the graduating class. For each class at each institution, the students came from a selection of different programs (as discussed below). In the university sample, only students in Engineering and Arts and Science faculties were included in the study, while in the college sample, students were taken from two- and three-year programs.

College

At the college, about 15% of the sample (over 100 observations) was dropped from the first-year class. This represented students from the health, interdisciplinary and trades programs. No students in the graduating

year who took the test came from these programs and so their exclusion among entering students was done in order to have consistency between the groups being compared.

Similarly, international students made up over 20% of the graduating class, but less than 2% of the sample in the entering cohort. International students are, therefore, still included in the samples, but are not separated out in the analysis.¹

A very small proportion of students was dropped due to unknown sex.

The total study sample includes 1,071 entering students and 484 graduating students. The sample of entering students is therefore more than twice as large as that of the graduating students.

University

At the university, a very small proportion of selected students (less than 5 observations) was dropped from the study due to unknown sex.

International students who would be returning to their home countries and therefore would not be available for the next phase of the project examining earnings through tax data were not invited by the university to take the test. Students unable to make it onto campus for various reasons (e.g., internship placement, exchange, distance education students) were also not invited.

The total study sample includes 1,075 entering students and 327 graduating students. The sample of entering students is therefore more than three times larger than that of the graduating class.

2.4 Variables Included in the Analysis

College

We divide the variables provided by the college based on their own administrative data and the data provided by ETS based on the HEIghten assessment into four groups: student characteristics, program, student assessments and time spent on the test.²

The student characteristics used in the analysis include sex, age at entry, first language, first-generation PSE and visible minority status.³ Age is broken down into six categories: students 18 and under, 19, 20–21, 22–23, 24–26 and 27 and above. First language is broken down into official language (i.e., English or French) and other. First-generation indicates that the student is from a family where neither parent completed any PSE,

1 When tested, international student outcomes were not statistically significant in any of models. This suggests that nothing related to international students and their scores is driving the results in any way.

2 A fifth group of variables reflecting student effort (described below for the university sample) could not be included in the analysis of college students since ETS removed the question related to this information from the assessment given to entering students.

3 A question regarding high school GPA was included in the HEIghten assessment but the data could not be included in the analysis because over 75% of students could not recall or did not answer.

as indicated in the college's administrative data. Visible minority status can either be yes, no, no response (student chose not to respond), or unknown (information is missing for unknown reasons). Students are grouped into four program categories based on classification criteria provided by the college. These are business; community and justice; engineering; and media and entertainment.

The college runs several assessments to gauge student performance upon entry. The main assessments of interest are for reading, writing and math. Not all students complete the assessments and math assessments in particular are not always required. For this study, based on a preliminary analysis of the relationships between the assessment scores and HEIghten scores, the raw assessment scores are divided into eight categories for the reading assessment and four categories for the writing and math assessments.

The college also obtained several time-based variables pertaining to the HEIghten test, including time spent on each test question. There are very strong correlations among these variables and we therefore use only the total amount of time spent on the entire test.

University

We divide the variables provided by the university (based on its own data and the data provided by ETS) for this study into three groups: student characteristics, faculty and student effort.

The student characteristics used in the analysis include sex, high school GPA, whether a student communicates better in English or another language and visible minority status. High school GPA is taken from the HEIghten survey and is recoded to include the following categories: GPA below 3.5, GPA 3.5–4, "I do not recall," "I prefer not to respond" and "unknown" (missing for other reasons). The question of whether a student communicates better in English or another language has possible responses of "No" (student communicates better in another language), "Yes" (student communicates better in English than any other language) and "Equal" (mastery of English and any other language is equal). In addition, HEIghten's race questionnaire is recoded into visible minority status with possible responses of "No" (not a visible minority), "Yes" (a visible minority), "No response" (the student chose not to respond) or "Unknown" (the information is missing for another reason).

Faculty is a binary variable and indicates whether a student is a part of the Faculty of Arts and Science or the Faculty of Engineering, the two faculties whose students were involved in the project.

The student effort variables come from the HEIghten test and include a binary measure of whether a student completed 75% of the test or more, as well as two effort-based survey questions. The first asks students whether they tried their best during the test, with the possible values being "No" (did not try their best), "Yes" (tried their best), or unknown (the information is missing). The second question asks students to assess the amount of effort they put into the test from "None" or "Little" (recoded due to sample size restrictions into one category: "Little or None"), "Moderate," "Lots," and "Best." In addition, an unknown category captures students for whom the response to this question is unavailable. Of course, this variable is self-reported and may not represent the true effort made and may be subject to various biases, such as students tending to respond with what they think are socially acceptable responses.

3. Methodology

3.1 Descriptive and Regression Analysis

The relationship between HEIghten test scores and various student characteristics, programs and measures of the effort the student put into the test are first analyzed in a descriptive framework using “violin graphs,” which show the full test-score distributions, as well as means, medians and quartile points, as described below.

Regression analysis is then used to relate HEIghten to the range of student and program characteristics available for the analysis in different combinations — first for the entering cohort, then for the graduating cohort, then for both cohorts combined — to test for whether the difference in the test scores between the cohorts differ across student and program characteristics.

The last point is addressed by running regression models of test scores including not only the student and program characteristics, but also interaction terms between these characteristics and the indicator for which cohort (class) students belong to. The focus of this particular analysis will be on the estimates of these interactions, which quantify by how much the difference in skill levels between cohorts change across various groups and provide statistical testing for whether these differences across groups are statistically significant.

3.2 Limitations

This section discusses some important limitations of this study. It is important to discuss these limitations here because they are pertinent to the interpretation of the results presented in the following sections, in particular comparisons across the two classes (cohorts) included in the analysis for each institution.

The limitations outlined here are related to sample selection issues and do not address whether the HEIghten test accurately measures critical-thinking skills. Discussion of any limitations of the HEIghten test are beyond the scope of this paper, where we simply assume that HEIghten indeed provides a satisfactory measure of students’ critical-thinking skills.

To start this discussion, it is important to emphasize that the data used in this study is cross-sectional. That is, the entering and graduating classes of students are not the same students observed at two different points in time, but rather two different sets of students who are captured at different points in their studies.

Comparisons within each cohort (or class) are, as a result, largely appropriate and meaningful, except for standard selection issues related to who takes the test in each cohort and how this might vary across the different student and program variables considered in the analysis (e.g., if test-taking rates differ by gender

or program of study and test taking is related to students' skill levels, this could bias comparisons across groups).

Comparisons across cohorts are, however, more problematic. First there is the issue of the attrition of students over time, such that not all those who start their studies are still there in the final year. If attrition is related to skill levels (which would not be surprising), the graduating cohort would tend to have higher scores (even if test participation rates were 100% for each class) simply because of these attrition effects and these effects may again vary across various student and program characteristics.

Secondly, test-taking rates may differ systematically for the two cohorts for a variety of reasons. For example, entering students may be more willing to participate in testing simply because they are new at the institution and generally more eager to participate in such activities, while students in their final year may be less likely to do so, perhaps because they are less generous with their time in terms of participating in such "extra" activities.

Third, to the extent the institutions recruit their entering and graduating cohort students into the test or otherwise conduct the tests differently for the two cohorts, this could further affect any cohort comparisons. At the university, the entering bachelor's degree students completed the test as part of a class assignment in a pre-selected set of courses in either the Engineering or Arts and Science faculty while the graduating students were simply invited to take the test and were financially compensated (with a \$25 gift card and entry to a \$750 prize draw) for their participation. At the college, the entering class took HEIghten as part of the standard suite of assessments conducted at the college while the graduating cohort was invited to participate in the assessment between classes and offered a \$40 incentive.

Finally, there may be underlying differences in the characteristics of the entering cohort and what the graduating cohort looked like when they started their programs, perhaps due to changes in recruitment or acceptance strategies on the part of the institution, due to any changes in the characteristics of the underlying potential student population (as was seen very clearly when the double cohort entered PSE), or any other such factors.

For all these reasons, observed differences in outcomes between the entering and graduating cohorts may be driven at least in part by differences in student characteristics related to the above selection factors.

Cross-institution comparisons would also be inappropriate. These are two very different institutions, in different PSE sectors (college versus university), with neither necessarily representative of the students or institutions in its sector. Other differences across the institutions pertaining more directly to the HEIghten test and related analysis include the students targeted with the assessment (students were drawn from different sets of faculties at the two institutions), the means of organizing the sessions (standard testing procedures versus more ad hoc arrangements established for the HEIghten test), attrition rates (affecting the graduating classes at the two institutions in particular) and the variables provided for the analysis.

4. Findings — College

4.1 Summary Descriptive Analysis

Figure 1 shows the distribution of the HEIghten test scores of the entering and graduating classes who took the test using what are known as kernel densities, which are essentially relatively simple mathematical functions fitted to best capture the underlying distributions. Figure 2 shows these distributions using a simple histogram instead, which naturally exhibits the same general shape but in a rougher way.

The overall distribution of test scores is almost identical for the two classes. These overall distributions show the full set of test scores rather than the simple means (159.3 for the entering class and 159.2 for the graduating class as shown in the summary statistics presented immediately below in Table 1) and show that beneath those comparable means are entire distributions of test scores that appear to be similar.

We return to the use of kernel functions of this type in the full descriptive analysis of the test scores, where full distributions of this type, along with a set of summary statistics (means as well as quartile cut points), are presented for each of the student characteristic, program and assessment score variables used in the analysis.

Table 1 reports i) the mean HEIghten scores for the entering and graduating cohort with respect to the full set of variables used in the analysis, including the student characteristics, program, the reading, writing and math assessments and the time spent on the test and ii) the distribution (as well as the actual counts) of students in the entering and graduating classes for each of the associated variables.

Figure 1: Kernel Density of HEIghten Scores by Cohort, College

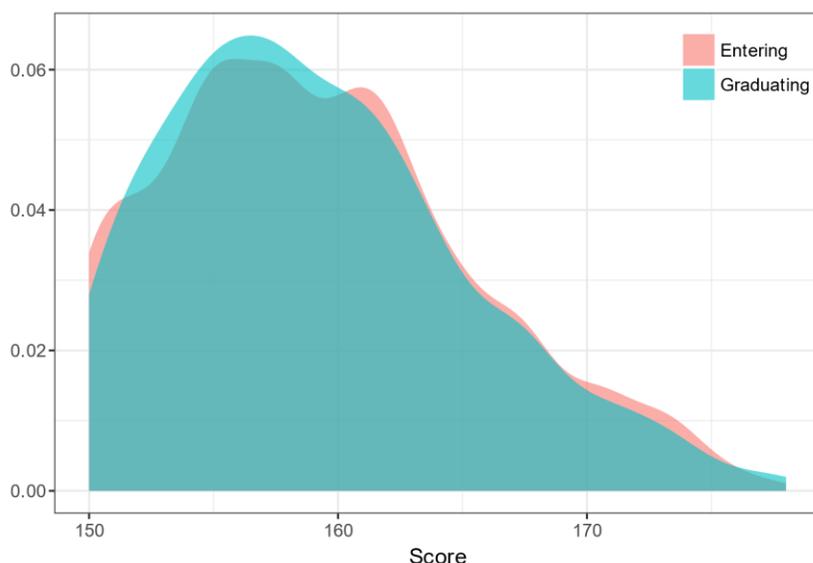
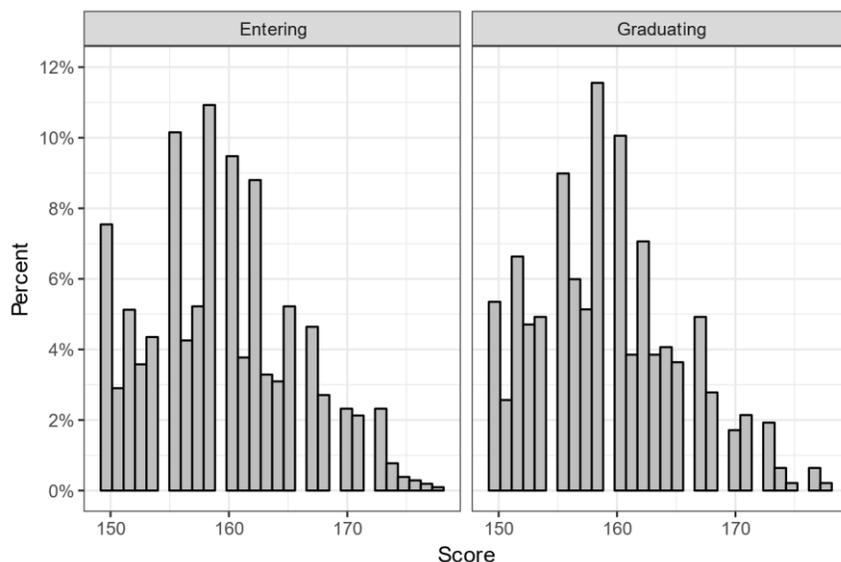


Figure 2: Distribution of HEIghten Scores by Cohort, College



At the overall level, the directions of the score patterns by various characteristics are generally similar between the entering and graduating classes, with the exception of the program differences. The mean HEIghten score is about 1.4 and 1.3 points higher for males for the entering and graduating students, respectively. The scores appear to increase with age, but the association does not seem to be very strong for either of the classes. Those whose first language is not one of the two official languages tend to score lower compared to those whose first language is an official language. Visible minority students score lower on average (2.5 and 3.8 for entering and graduating, respectively) compared to those who are not visible minorities. Engineering students score the highest on average within the entering cohort, whereas media and entertainment students have the highest mean score for the graduating cohort. For both classes, assessment scores are positively related to the HEIghten scores.

However, these patterns are to some extent influenced by how these student, program, assessment score and time variables are related to each other (e.g., males are more concentrated in engineering programs).⁴ Therefore, we use a regression model approach in Section 4.3 to examine the score patterns by various characteristics taking into account differences in other factors.

Comparing the classes in terms of HEIghten scores, the differences between the two classes (entering and graduating) appear to be relatively small for each of the student, program, or assessment score variables. However, we remind the reader that these represent different sets of students and there could be a range of potential biases involved in such comparisons (as described in the Introduction). Table 1 points to just how different the two samples are. The p-value in the “Diff.” column indicates whether the two samples are significantly different with respect to the sample shares of each of the associated variables, with a p-value of

⁴ There could also be unobserved differences between the two classes that might also contribute to the score difference.

.05 or less indicating a difference between the two classes at the 5% significance level and p-values between .05 and .1 indicating significance at the 10% level.

The differences in the overall sample shares are in fact statistically significant for the great majority of variables. These include the majority of student characteristics, as well as program.

Again, we use regression models to look at the differences in HEIghten scores between the entering class and graduating class after taking into account differences in characteristics between the two cohorts.⁵

Table 1: Descriptive Statistics, College

		Distribution					Mean HEIghten Scores		
		Entering		Graduating		Diff.	Entering	Graduating	Diff.
		%	Count	%	Count	p-value	Score	Score	p-value
All		100	1,071	100	484		159.3	159.2	0.68
Sex	Female	42.7	457	54.3	263	0.00	158.5	158.6	0.90
	Male	57.3	614	45.7	221	0.00	159.9	159.9	0.97
Age Group	<=18	34.5	370	20.2	98	0.00	158.5	158.4	0.91
	19	21.0	225	17.1	83	0.07	159.0	157.6	0.06
	20–21	18.4	197	19.6	95	0.57	160.0	159.4	0.36
	22–23	9.52	102	12.8	62	0.06	159.6	162.3	0.01
	24–26	6.63	71	10.3	50	0.02	161.4	158.2	0.01
	27+	9.90	106	19.8	96	0.00	159.9	159.6	0.75
Mother Tongue	First language official	92.4	990	70.9	343	0.00	159.6	160.2	0.10
	First language other	7.56	81	29.1	141	0.00	155.7	156.6	0.22
First Gen. PSE	No	73.8	790	75.6	366	0.43	159.5	159.3	0.52
	Yes	26.2	281	24.4	118	0.43	158.7	158.8	0.81
Visible Minority	No	56.2	602	34.5	167	0.00	160.1	161.4	0.02
	No response	3.74	40	3.72	18	0.99	157.9	156.6	0.38
	Yes	29.2	313	36.0	174	0.01	157.6	157.6	0.92
	Unknown	10.8	116	25.8	125	0.00	160.3	158.7	0.04
Program	Business	24.8	266	36.2	175	0.00	158.6	158.8	0.71
	Community and justice	31.4	336	31.2	151	0.95	158.8	158.0	0.14
	Engineering	34.5	369	28.3	137	0.01	160.2	160.4	0.80
	Media and entertainment	9.34	100	4.34	21	0.00	159.7	163.0	0.04
Assess Reading	1	10.9	117	6.20	30	0.00	154.8	154.5	0.62
	2	10.5	112	5.37	26	0.00	155.3	156.6	0.21
	3	11.9	127	6.41	31	0.00	157.3	158.5	0.24

⁵ We also allow for the differences in scores between the classes to vary by student and program characteristics, assessment scores and time spent on the test.

		Distribution					Mean HEIghten Scores		
		Entering		Graduating		Diff.	Entering	Graduating	Diff.
		%	Count	%	Count	p-value	Score	Score	p-value
	4	11.6	124	9.92	48	0.32	157.8	158.3	0.48
	5	12.0	128	8.47	41	0.03	159.5	159.2	0.79
	6	11.6	124	7.23	35	0.00	160.4	161.7	0.17
	7	12.4	133	9.09	44	0.04	163.1	162.5	0.49
	8	13.4	144	10.3	50	0.07	166.1	166.7	0.56
	(Missing)	5.79	62	37.0	179	0.00	155.8	157.2	0.06
Assess	1	27.4	293	13.8	67	0.00	157.2	158.2	0.20
Writing	2	41.7	447	24.0	116	0.00	159.4	159.9	0.43
	3	18.6	199	17.4	84	0.56	161.9	161.3	0.46
	4	6.54	70	6.61	32	0.96	163.0	162.8	0.90
	(Missing)	5.79	62	38.2	185	0.00	155.6	157.4	0.02
Assess Math	1	11.6	124	8.06	39	0.03	156.7	156.0	0.35
	2	13.9	149	8.26	40	0.00	158.3	159.9	0.12
	3	13.2	141	12.8	62	0.85	160.7	161.3	0.50
	4	11.4	122	17.8	86	0.00	162.7	159.3	0.00
	(Missing)	50.0	535	53.1	257	0.25	159.0	159.0	0.88
Time Spent	Mean	35.9		37.5		0.00			

4.2 Full Descriptive Analysis

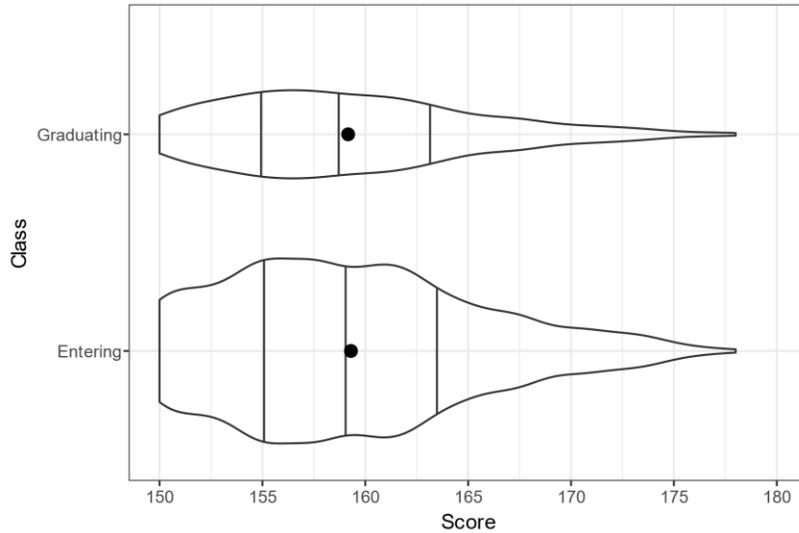
Test Scores by Class

All subsequent figures present results using what are known as violin plots. We use this approach because they show the (mirrored) kernel densities as explained above (the outline of the plot), the size of the sample (the thickness of the plot), quartiles (vertical lines from left to right: 25th, median, 75th percentiles) and the sample mean (using a dot). In other words, it provides a very comprehensive presentation of the HEIghten test scores.

Figure 3, for instance, shows the violin plots for the entering and graduating classes. First, what was just discussed about Figure 1 in the main text can also be seen in Figure 3. In fact, the violin plots merely replicate the kernel densities seen above and mirror them to form the outline of each plot. In other words, the bottom part of the outline of the plot simply mirrors the top part of the outline. As the kernel distributions show, the test score distributions appear to be similar for the two classes.

The median and mean scores for entering and graduating classes are the same or almost identical at 158 for medians and around 159.2–159.3 for means.

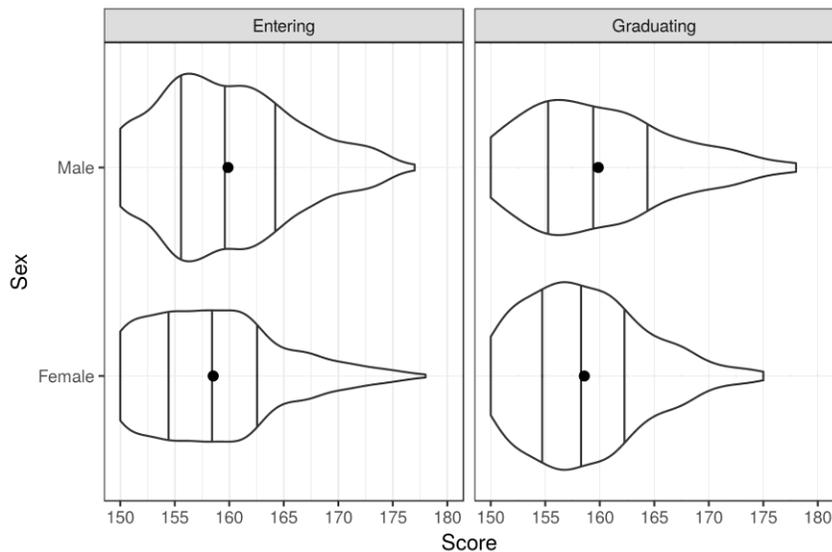
Figure 3: Violin Plots of HEIghten Scores by Cohort, College



Test Scores by Selected Characteristics

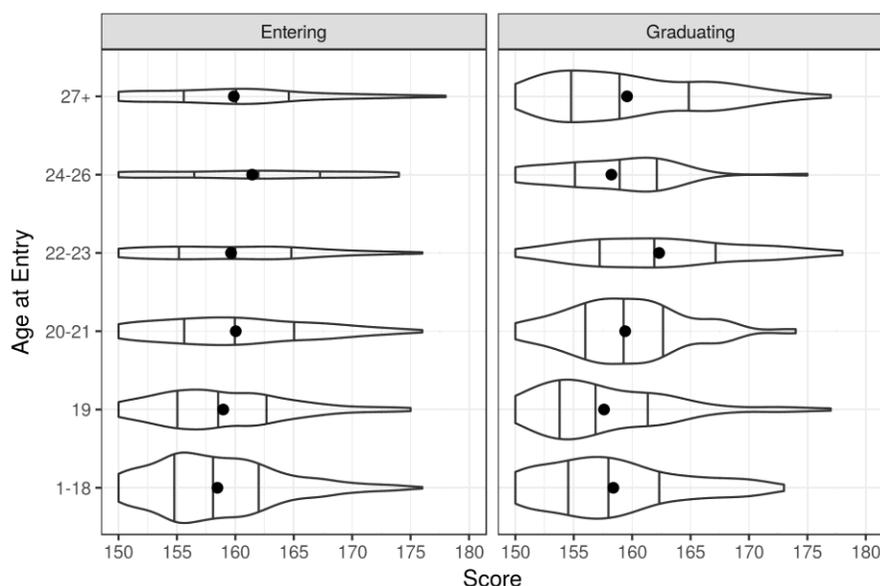
Turning our attention to differences by sex, we note that the males constitute a slight majority of the overall students in the entering class; the reverse is true among the students in the graduating class. The overall performance of both males and females appears similar, although the upper quartile of scores for males is higher in both the entering and graduating class. The distribution of high scores appears to taper off faster for females than males, especially in the graduating class.

Figure 4: Violin Plots of HEIghten Scores by Cohort and Sex, College



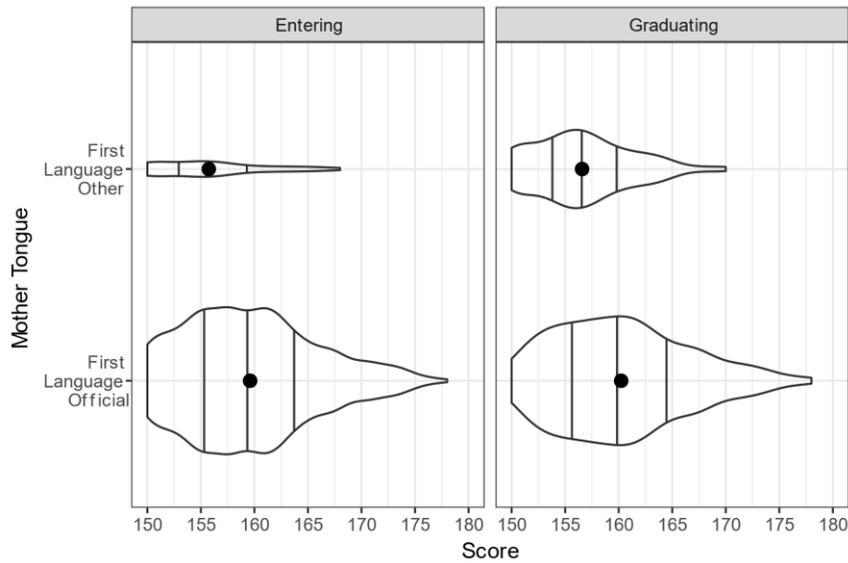
Results by age at entry show that the two entering and graduating classes appear to have some differences in the age distribution. The graduating cohort is older at entry, with relatively more students present at ages 22 and up compared to the entering class. Among the students in the entering class, the overall score distributions appear to be similar, although students 19–27 have slightly higher scores on the test. There does not seem to be a clear pattern in the score distributions for the graduating class. Students 18 and under tend to do slightly better than 19-year-olds. Students aged 22–23 perform the best, while students 24–26 are closer to those 18 and under in performance. Students aged 27 and above tend to score close to 24–26 year olds, but have a slightly higher mean and upper quartile with the right tail of distribution tapering off relatively slowly.

Figure 5: Violin Plots of HEIghten Scores by Cohort and Age, College



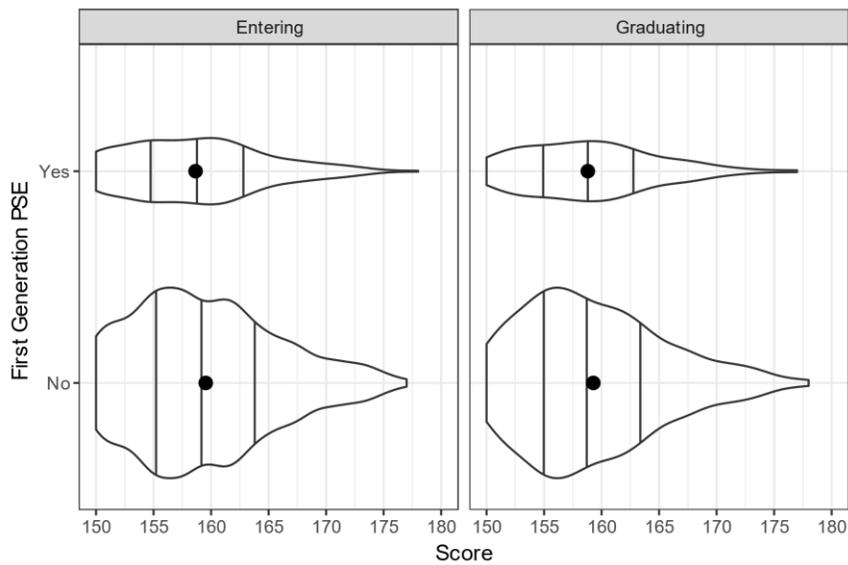
Most of the students in both the entering and graduating classes speak an official language (either French or English) as their first language. This is especially true in the entering class, where the number of students who speak a different first language is very small. In both cases, students whose first language is not an official language do not score as well. In both cases, the score of the upper quartile of those with another first language is close to the median of students who speak an official language as their first language.

Figure 6: Violin Plots of HEIghten Scores by Cohort and Language, College



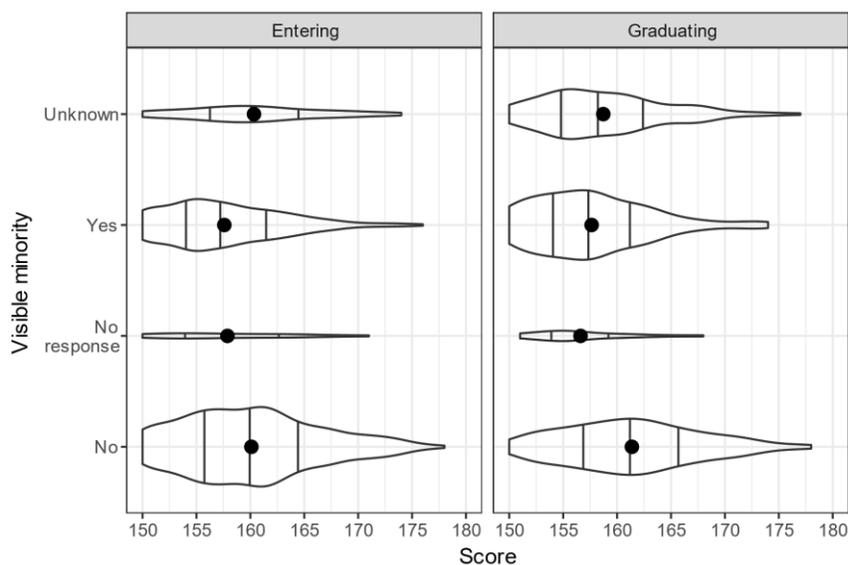
There is little difference in the distribution of scores of students based on the first-generation status. Most students are not first-generation in both samples. Students who are not first-generation postsecondary students have only a slightly higher mean score and upper quartile and a slightly higher distribution of scores at the right tail.

Figure 7: Violin Plots of HEIghten Scores by Cohort and First Generation Status, College



The sample of visible minorities is relatively larger in the graduating than the entering class. Students who are not visible minorities appear more common in the entering than the graduating class, although this may be the result of the relatively higher number of students for whom this status is unknown in the graduating class. Students who preferred not to respond to this question were very few in both classes. The pattern is the same for both classes: visible minorities had lower mean, median and upper quartile scores compared to other categories for this variable, except for those who did not give a response.

Figure 8: Violin Plots of HEIghten Scores by Cohort and Visible Minority Status, College



Score distributions by program grouping appear to be similar for students who begin their program in either business or community and justice, in both the entering and graduating classes. In comparison, engineering students have higher scores across the board and this is especially true for the graduating class. Media and entertainment students scored comparably to the business students in the entering class, and better in the graduating class; however, these students constitute a relatively small part of the graduating class.

The college reading assessment score appears to show a positive association with test performance. This is the case for both the entering and graduating students, although there are relatively more students with missing information for this test in the graduating class. Nonetheless, the pattern is that across the scoring scale, those with strong reading test scores tend to score better on HEIghten.

As with the reading assessment score, there is a positive association between the writing assessment score and test score. Again, there are relatively more students for whom the writing score is missing in the graduating class.

Figure 9: Violin Plots of HEIghten Scores by Cohort and Program, College

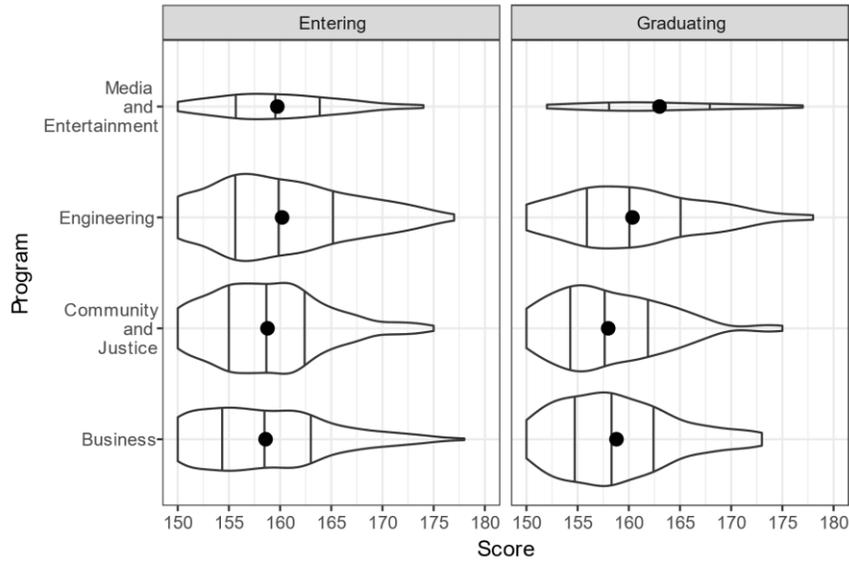


Figure 10: Violin Plots of HEIghten Scores by Cohort and Reading Assessment Score, College

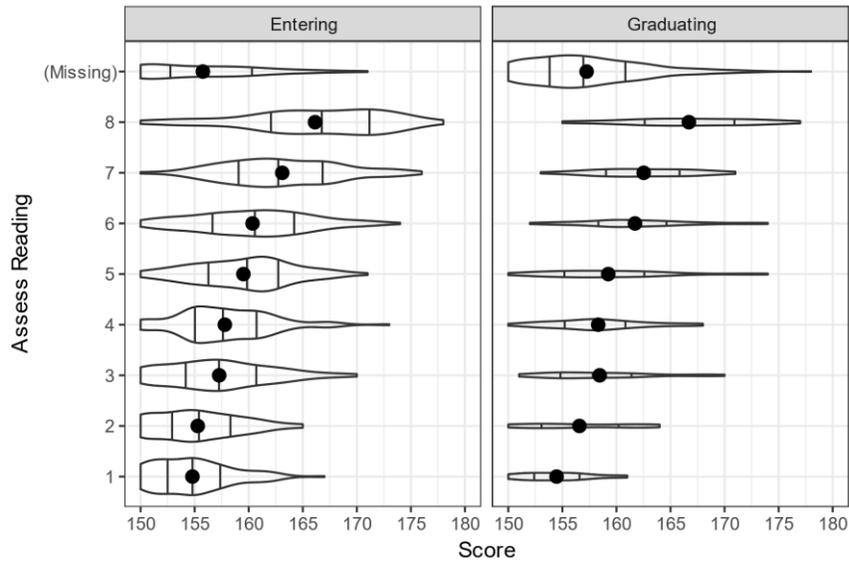
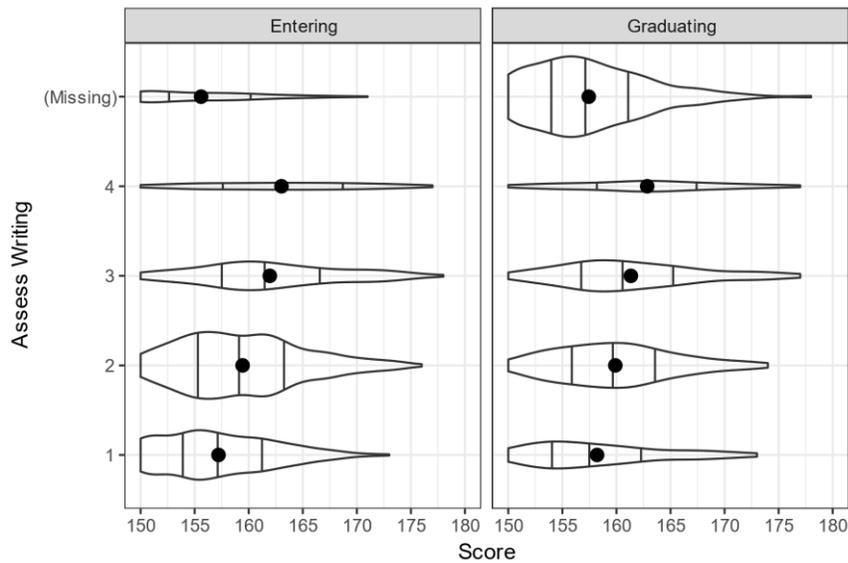
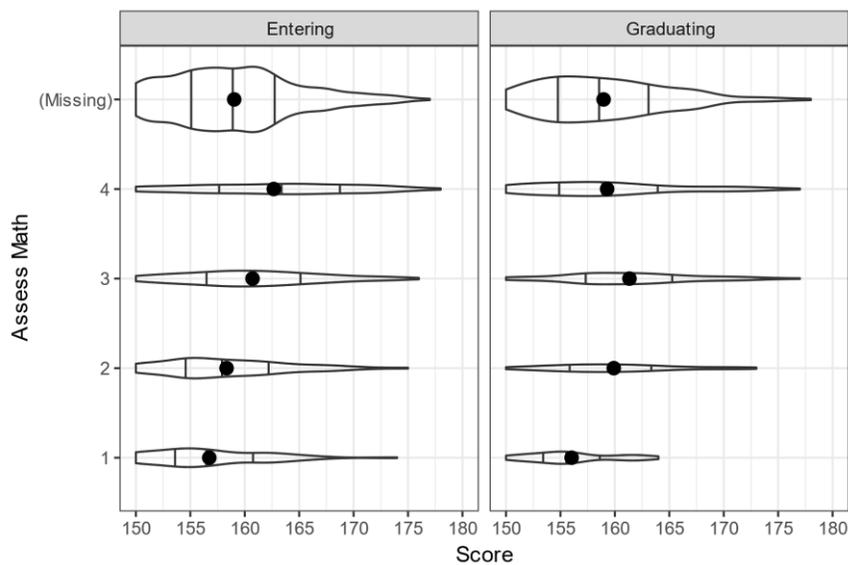


Figure 11: Violin Plots of HEIghten Scores by Cohort and Writing Assessment Score, College



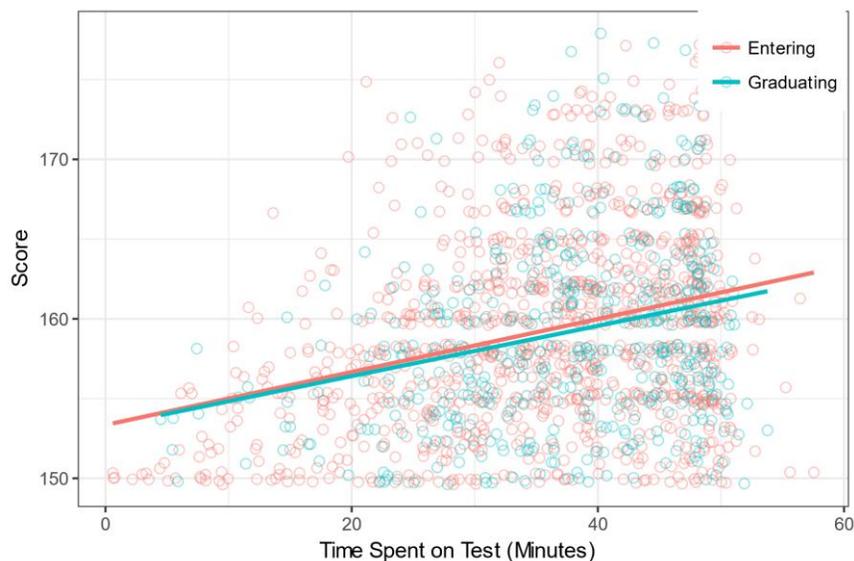
Next, we analyse the math assessment score. This assessment is not required for a large subset of students, so the missing category in this case is large for both entering and graduating students. For those students for whom this information is available, there is some evidence that higher math assessment scores are associated with higher HEIghten scores, although the pattern is less clear than it is for reading and writing assessments. The one exception is found in the graduating sample where students who score 3 out of 4 on our scale actually do slightly better than those who score 4 out of 4.

Figure 12: Violin Plots of HEIghten Scores by Cohort and Math Assessment Score, College



Finally, we take note of the time students spent on the test and demonstrate the result in a scatterplot with two linear regressions (one for the entering and one for the graduating class) overlaid to illustrate the pattern. Overall, and perhaps unsurprisingly, students who spent more time on the test generally did better and the overall slope for this variable is similar for the entering and graduating class.

Figure 13: Scatter Plot of HEIghten Scores and Time Spent on the Test by Cohort, College



4.3 Regression Analysis

Entering Cohort

We now turn to the results of the regression analysis of HEIghten test scores, which are included as a separate section at the back of the paper, starting with the college results followed by the university results.

Table R1 shows the results for four different models for the college students in the entering class. The first is a model that includes student characteristics: sex, age, first language and whether the student is first-generation PSE or identifies as a visible minority. We then show the results for a second model, which includes all the controls from the student characteristics regression, but also the student's program. The third table further adds the scores from the reading, writing and math assessments (administered to students at entry) to the model. The final model adds the total time spent on the test (in minutes). These models are all run with 1,071 observations.

The purpose of running these different models is to understand how the student characteristics are related to HEIghten scores both when taken as a group in the first model (i.e., all student characteristics are included in model 1), and then after taking account of the additional variables included in the subsequent models. The models also, of course, identify the relationship between HEIghten scores and the additional

sets of variables added in the second, third and fourth models (i.e., program, assessment scores, time spent on the test). Taken together, these different sets of results provide a full picture of how HEIghten scores are related to the different sets of variables included in the analysis.

In almost all cases, the explanatory variables included in the models represent categorical variables, the only exception being time spent on the HEIghten test. This means that the coefficients shown in the tables represent the estimated differences in test scores, on average, associated with each variable (category) included in the model in comparison to the related omitted (or baseline) category (e.g., students who were older at entry in comparison to the youngest group) while taking the other variables included in the model into account.

When looking at the student characteristics only, we observe that males score about 1.7 points (which is close to the 1.4 point-difference seen in Table 1) higher than females when all other student characteristics are taken into account. This difference decreases slightly when program variables are added and then decreases further to about 0.7 when assessment variables are added.

It is not surprising that the estimated gender gap decreases slightly when program is added, indicating that some of the male effect was capturing an engineering effect: that is, males tend to be relatively more concentrated in engineering programs and students in engineering programs tend to have higher scores than others (see the table and the related discussion below). Following a similar logic, the more substantial decrease in the male advantage when the assessments are included can be explained by the fact that males tend to score higher on the assessments, which tend to be positively associated with HEIghten scores. Including time spent in the model does not affect the estimated gender gap in mean scores.

The effect of age at entry appears to be important and scores generally rise with age, but with some variation around this tendency in the first model (i.e., student characteristics only). On average, 19-year-olds score slightly higher than the omitted group (those 18 and under), but that difference is not statistically significant. Students aged 20–21 score about 1.9 points higher (in this first model), while 22–23-year-olds score 1.7 points higher. Students aged 24–26 score the best, almost 3.8 points higher than the omitted group, while those aged 27 and above score about 2.6 points higher. These patterns stay more or less the same when program variables are included, but they disappear when assessment variables are added to the model (third model), except for the 24–26 age group, suggesting that age is related to the assessment scores. Adding time spent to the model does not change the coefficients for the different age groups.

For convenience, we use the word “effect” to represent the relationship between the explanatory variables included in the models and HEIghten scores, but this should not be interpreted as necessarily representing a causal influence.

As seen in Table 1 (above), those with a first language other than English or French score about 4 points lower than those without one of the official languages. This difference is smaller, at about 3.2 points, when student characteristics are added to the model (model 1 in Table R1) and even smaller, at about 1.6 points, when all variables are included in the model. This decrease in language effects across the different models suggests that they are explained, statistically speaking, at least in part by differences in student characteristics, assessment scores and time spent on the test.

These language differences could also be related to those not having English or French as their first language experiencing difficulties with the test rather than reflecting their actual levels of critical thinking, but that is not something that can be tested with the data available.

First-generation students are also at a slight disadvantage, scoring about one point lower than others, holding other student characteristics constant. However, this difference disappears altogether when assessment scores are added, meaning that the difference in scores between first-generation students and others (seen at the overall level in Table 1) is related to differences in their assessment scores.

Students who identify as a visible minority score about 2 points lower on the test than those who do not, while those who offered no response to this question score about 2.3 points lower (model 1). Students for whom this information is missing for some other reason did about the same as the baseline group (i.e., those who are not visible minorities). These differences across groups remain when program variables are added, but they are smaller (at about 1.1 points) when the assessments are added. Adding time spent on the test slightly increases the estimated difference between those who identify as a visible minority and those who do not.

We now move from the student characteristics to the other variables included in models 2, 3 and 4. In model 2, which adds program, we see that all programs have, on average, higher scores than the business students, but only those in engineering show a statistically significant difference (1.7 points). This difference decreases to about 0.8 points when all variables are added to the model, suggesting that it is to some extent explained by how students scored in their assessment tests and, to a lesser degree, the time spent on the test.

Looking at model 3, where assessment scores are added to the model, we see that students with higher assessed reading scores (at entry to the college) do far better than those with lower assessment scores (model 3). For example, those in categories 6, 7 or 8 score 4.3, 7.1 and 9.5 points higher, respectively, than those who scored in the lowest category (the omitted group), while those with a missing reading assessment also do considerably better than the low scoring group, by about 4.9 points, on average. Adding time spent to this model (model 4) decreases the magnitude of these differences a bit, but the generally strong association between reading assessments and HEIghten scores remains.

For the writing assessment, those who score in categories 3 and 4 (i.e., third and fourth quartiles) attain higher scores, on average, than those in the bottom level by 1.8 and 1.4 points, respectively. Students for whom a writing score is not observed do 3.2 points worse than the comparison group (bottom quartile). These patterns remain more or less the same when time spent on the test is added to the model, with the most notable changes being an increase, to 1.7 points, in the difference between fourth and first quartiles, and a decrease, to 2.1 points, in the difference between those who scored within the bottom quartile and those whose score is missing.

For the math assessments, only students in the highest category score higher (2.2 points on average), than those in the omitted (lowest) group. This result remains when time spent is added to the model, with only a slight decrease in the difference in scores by math assessment.

Finally, model 4 shows that time spent on the HEIghten test appears to be an important factor, with each additional minute spent on the test corresponding to a 0.14 point increase in scores.

In sum, many of the differences in HEIghten scores across the student characteristics identified in the first model, which includes only those student characteristics, remain more or less the same when the program variables are added to the model, but adding the assessment scores to the model generally decreases the estimated differences, while adding time spent on the test to the models generally has little additional effect on the estimated differences. A substantial part of the differences in HEIghten scores across different groups of students are, therefore, related to assessment scores; that is, those who show higher HEIghten scores also tend to score better on the assessment tests.

These findings could be thought of, first of all, as a kind of informal general validation of the HEIghten test, as seen in how HEIghten scores are related to the assessment scores and how the assessment scores are related to the observed differences across student groups. And secondly, they help us understand the patterns of HEIghten scores across student characteristics in terms of i) what the overall differences are (Table 1 above and model 1 here) and ii) what differences remain after taking the other variables included in the models into account (models 2–4).

Graduating Cohort

We ran the same set of regressions for the graduating cohort as we did for the entering class: characteristics, program, assessment and time spent on the test.

Looking at the student characteristics model (model 1), males score a statistically significant 1.6 points higher, on average, than females on the HEIghten test, controlling for other student characteristics. As previously seen for the entering students, this difference decreases when program is added to the model, and decreases further when the assessment scores are also included, which continues to hold when time spent on the test is added as well.

Students who were age 18 or under at entry and those who were age 19 score about the same, while students who were 20 or older tend to score higher. For example, those who were 20–21 years old score about 1.24 points higher compared to those who are 18 and under, although this difference is not statistically significant. Students who were 22–23 years old score a statistically significant 3.7 points higher on the test, while those 27 and above score 2 points higher compared to those who are 18 and under (again significant). The higher scores for the 22–23 and 27 and above groups persist when the program variables are added, but these differences decrease for the 22–23 group and disappear for the 27 and above group when the assessments and time spent on the test are added (models 3 and 4).

Students with a non-official first language score about 2.5 points lower in the first two models, but this difference decreases (to about 1.4–1.8 points) when the assessment and time spent variables are added.

Unlike with the entering students, first-generation status is not statistically significant except in the model including student characteristics.

Students who are visible minorities score about 2.5 points lower than those who are not visible minorities; those who refused to respond to this question score 3.8 points lower, while students for whom this information is otherwise missing score about 1.9 points lower than non-minorities. These differences are generally the same when program is taken into account, but, again like the entering class, diminish when assessment scores are included in the model, indicating that minority status is related to the assessment scores (i.e., minorities tend to score lower on the assessment tests, which would be associated with lower HEIghten scores). Adding time to the model then increases the differences slightly, but the patterns remain the same.

In model 2, where program is added, community and justice students are seen to score a statistically significant 1.6 points lower, on average, than business students. In comparison, media and entertainment students score higher than business students, by a significant 3 points. Engineering students score slightly higher (1.1 points) than business students. The media and entertainment and engineering differences (compared to business) get smaller and are no longer statistically significant once the assessment and time variables are added, while the community and justice difference slightly increases.

Again, as seen for the entering students, there appears to be a strong relationship between the reading assessment and HEIghten scores. Those who score higher on the reading assessment also tend to score higher on HEIghten, which is seen in both models 3 and 4. Unlike the entering students, those who have no reading assessment scores do about as well on the HEIghten as those who score in the lowest reading assessment category.

In contrast, HEIghten scores do not appear to be related to writing scores. One exception is that those who did not take the test or for whom the information is missing score about 3 points higher than those who score in the lowest writing assessment category.

HEIghten scores are also at least somewhat related to the math assessment scores. Students with math assessment scores in the second and third categories score about 3 points higher and those whose math score is in the fourth category score about 1.5 points higher compared to the lowest category.

Time spent on the test is again significantly related to HEIghten scores. Those who spend an extra minute score, on average, 0.14 points higher, taking all other characteristics into account, which is the same result we observe with the entering class.

One of the take-away messages for the graduating class is that there appear to be some differences in HEIghten scores by age, language, visible minority status, program, reading and math assessment scores. The gender gap in HEIghten scores is to a substantial extent related to differences in the programs in which males and females are enrolled and is also related to the reading, writing and math assessments administered to students at entry.

Beyond that, differences across most of the other student characteristics are related at least in part to assessment scores, meaning that student characteristics are correlated with assessment scores, which have their own effect on HEIghten test scores.

Finally, these patterns across student characteristics differ for the entering and graduating classes, which suggests that there are associated differences in *the changes* in HEIghten scores between the entering and graduating classes across student characteristics as well as the other variables included in the models. We test for these differences more formally by estimating regression models that include a complete set of interactions with the class the student belongs to, reported below.

Cross-cohort Comparison

Table R3 shows the results from a set of pooled regression models which include students from both the entering and graduating classes. All of the explanatory variables outlined in the previous section are again included in these models. However, we also add a set of “class” interactions.

First, an overall class indicator (“graduating”) represents a general indicator of whether the student is in the graduating class as opposed to the entering class. Then, associated with each set of variables included in the models (male, age at entry, etc.), a variable representing the interaction of the (graduating) class indicator and each of the related individual variables in question is included (“graduating * male,” “graduating * 19,” etc.).

Table R3 presents the results from the regression models that add class interactions to the variables included in each of the models, which follow the same progression as in the separate regressions by class presented above. A very simple model that includes only the graduating class indicator is shown first (model 1), only to demonstrate that the overall indicator represents the overall difference in scores between the two classes (-0.14). Then the class variable is interacted with the student characteristic variables (model 2), then class interactions with program variables are added (model 3), then with the assessment scores (model 4) and finally, interactions are added to the effort-based variables. These models allow the difference in scores between classes to vary across student and program characteristics.

While this full set of results is presented for completeness, we focus our discussion on the last model, which includes the full set of explanatory variables and focus further on the interaction terms of interest. These identify the patterns across groups of students in the differences in skills between the entering and graduating class.

Once we add the class interactions, we see that the differences in scores between the graduating and entering classes do indeed differ across some of the characteristics/variables represented in the models. For example, the coefficient of the “graduating * 22–23” is 1.84, which means that the difference between the two classes is 1.84 points higher for those who are 22–23 than for those who are 18 and below.

The other significant differences are seen by program and assessment scores. The difference in scores between the two classes is 2.28 points lower for community and justice students than for the comparison business program, holding all other factors constant; larger for those who scored within the second and sixth category of the reading assessment score; about 2 points smaller for those who scored in the second and third quartiles of the writing assessment score distribution; and about 3 points for those who scored in the second and third quartiles of the math score distribution compared to those who scored in the bottom category or quartile. Moreover, those who miss their writing scores also seem to have larger (by about 5 points) difference in skill levels between graduating and entering classes compared to those who scored within the lowest category.

5. Findings – University

5.1 Summary Descriptive Analysis

As with the college results, the presentation of university results begins with a simple descriptive analysis, starting with the kernel density and histogram representing the overall distribution of the test scores.

Figure 14: Kernel Density of HEIghten Scores by Cohort, University

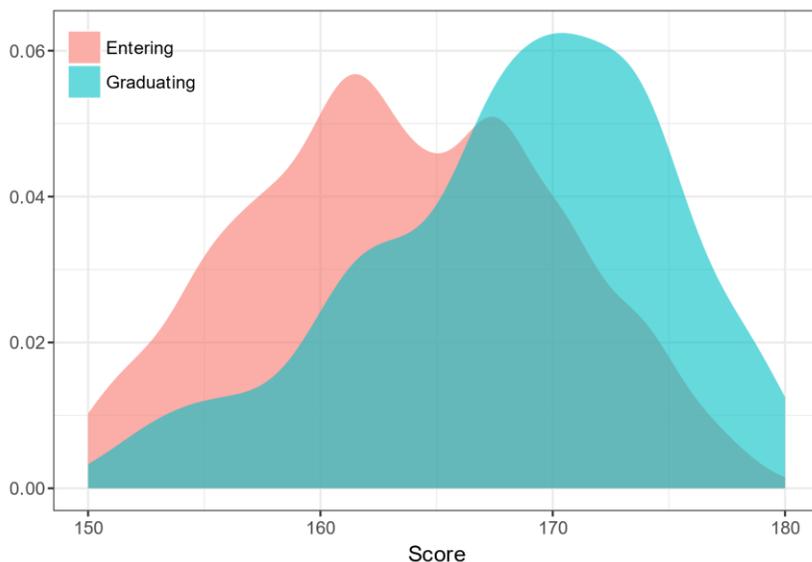
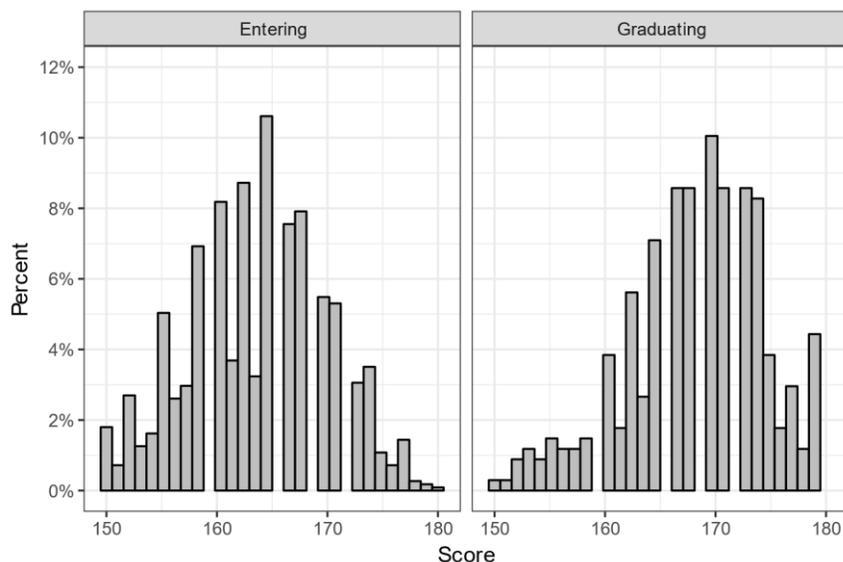


Figure 15: Distribution of HEIghten Scores by Cohort, University



The entering class has a peak distribution around a score of 162 and a second and higher concentration around a score of 167. In contrast, the distribution of test scores for the graduating class is shifted to the right, with the greatest cluster of scores occurring around the 170 mark. These distributions are consistent with the simple mean scores (163.5 for the entering class and 168.3 for the graduating class as also shown in Table 2 below), but capture the differences in test scores between the two classes in much more detail.

Table 2 shows a similar set of descriptive statistics as shown for the college students earlier, thus showing the mean HEIghten scores for each class for each variable, as well as the distributions of the samples with respect to each of the different sets of variables included in the analysis.

The mean HEIghten score is about 2 points higher for males for the entering students, while the gender difference for the graduating students is very small. For both classes, high school GPA appears to be positively related to the HEIghten score. Whether a student communicates better in English is also related to the performance on the HEIghten test for the entering students, but not for the graduating students. For both classes, visible minorities tend to do worse on the test compared to those who are not a visible minority. Students from the Faculty of Engineering tend to score higher than students from the Faculty of Arts and Science in the entering class, whereas the two faculties show very similar scores for the graduating class. The number of questions answered on the test and other effort-related variables appear to be positively associated with the HEIghten score for both classes.

As was done for the college students, we again use regression models in Section 5.3 to examine how HEIghten scores patterned across the variables included in the analysis while taking account of the other factors, as opposed to the simple differences just reported.

Comparing the two classes in terms of HEIghten scores, there appears to be significant differences in mean scores between the two classes for almost all of the student characteristics, faculty and effort-related variables. Again, we emphasize that these are different sets of students and as a result there could be potential biases in such comparisons. Table 2 confirms this by showing that the two samples are in fact quite different. One of the most interesting findings here is that the proportion of students with an incoming (high school) grade of 3.5–4 (i.e., the highest category) goes from 52.3% of the sample for the entering class to 74.3% for the graduating class. The share of the graduating class made up of graduates from the Faculty of Arts and Science also rises, while the share from the Faculty of Engineering declines commensurately. Finally, the graduating class put more effort into the test.

We would expect overall scores to be affected by these sample differences, presumably driving the scores of the graduating class higher due to these sample effects alone — such as the observed differences in incoming grades, since those with higher grades tend to score higher on HEIghten.⁶

As done for the college students, we use regression models that include both classes in the sample in the last part of Section 5.3 to examine the differences in scores between entering and graduating students after taking into account differences in characteristics between the two cohorts.

Table 2: Descriptive Statistics, University

		Distribution					Mean HEIghten Scores		
		Entering Class		Graduating Class		Diff.	Entering	Graduating	Diff.
		%	Count	%	Count	p-value	Score	Score	p-value
All		100	1,075	100	327		163.5	168.3	0.00
Sex	Female	63.3	680	68.5	224	0.08	162.8	168.2	0.00
	Male	36.7	395	31.5	103	0.08	164.8	168.5	0.00
High School GPA	< 3.50	10.7	115	9.8	32	0.63	160.9	166.0	0.00
	3.50–4.00	52.3	562	74.3	243	0.00	164.7	169.0	0.00
	I do not recall	12.3	132	7.0	23	0.00	162.8	166.2	0.03
	I prefer not to respond	3.3	35	0.9	3	0.00	160.3	160.3	0.99
	Unknown	21.5	231	8.0	26	0.00	162.8	167.7	0.00
Communicates better in English	No	23.8	256	27.2	89	0.22	162.7	168.8	0.00
	Yes	67.3	724	63.9	209	0.26	163.8	168.2	0.00
	Equal	8.8	95	8.9	29	0.99	163.7	167.8	0.00
Visible Minority	No	62.9	676	59.0	193	0.21	164.4	169.2	0.00
	No response	1.0	11	1.8	6	0.31	160.1	165.0	0.21
	Yes	22.1	238	34.3	112	0.00	161.7	166.8	0.00
	Unknown	14.0	150	4.9	16	0.00	162.4	168.8	0.00

⁶ Again, we point to the likelihood that the samples are different in other ways not captured by the variables we observe, thus affecting the comparisons across the classes in ways that cannot be predicted.

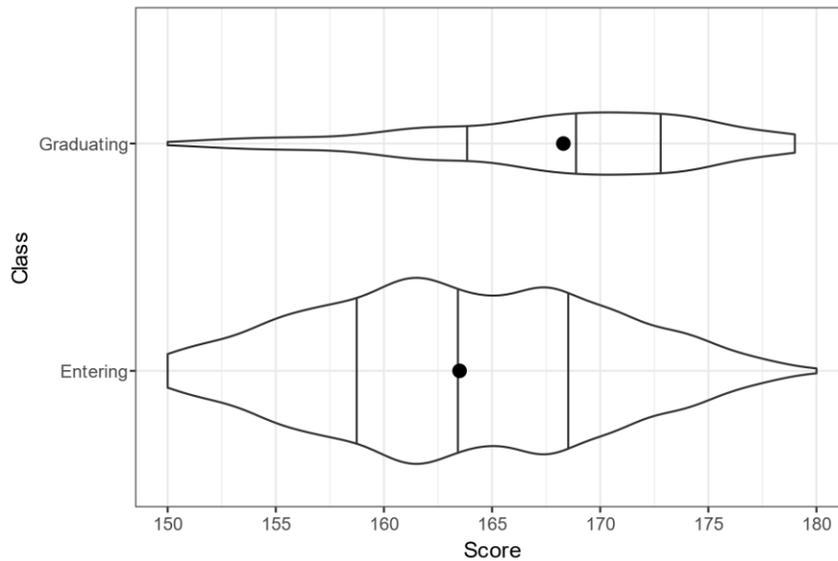
		Distribution					Mean HEIghten Scores		
		Entering Class		Graduating Class		Diff.	Entering	Graduating	Diff.
		%	Count	%	Count	p-value	Score	Score	p-value
Faculty	Arts and Science	61.6	662	71.9	235	0.00	162.0	168.3	0.00
	Engineering	38.4	413	28.1	92	0.00	165.8	168.2	0.00
Answered ≥75% of test	Yes	4.5	48	2.8	9	0.12	155.8	161.3	0.05
	No	95.5	1,027	97.2	318	0.12	163.9	168.5	0.00
Tried their best during test	No	24.0	258	14.1	46	0.00	160.8	163.1	0.03
	Unknown	46.0	494	18.7	61	0.00	165.1	168.3	0.00
	Yes	30.0	323	67.3	220	0.00	163.2	169.4	0.00
Effort put into the test	Little or None	16.0	172	8.0	26	0.00	159.8	163.2	0.04
	Moderate	49.7	534	44.6	146	0.11	163.5	167.9	0.00
	Lots	20.7	223	31.2	102	0.00	165.5	169.5	0.00
	Best	8.2	88	13.8	45	0.01	167.2	169.9	0.03
	Unknown	5.4	58	2.5	8	0.01	161.2	168.1	0.00

5.2 Full Descriptive Analysis

Test Scores by Class

The median and the mean scores are almost identical for the entering class, both around 163. In comparison, the median and mean scores of the graduating class are higher at 170 and 168, respectively. Therefore, the graduating class scores around 5 points higher on average on the HEIghten test than the entering class.

Figure 16: Violin Plots of HEIghten Scores by Cohort, University



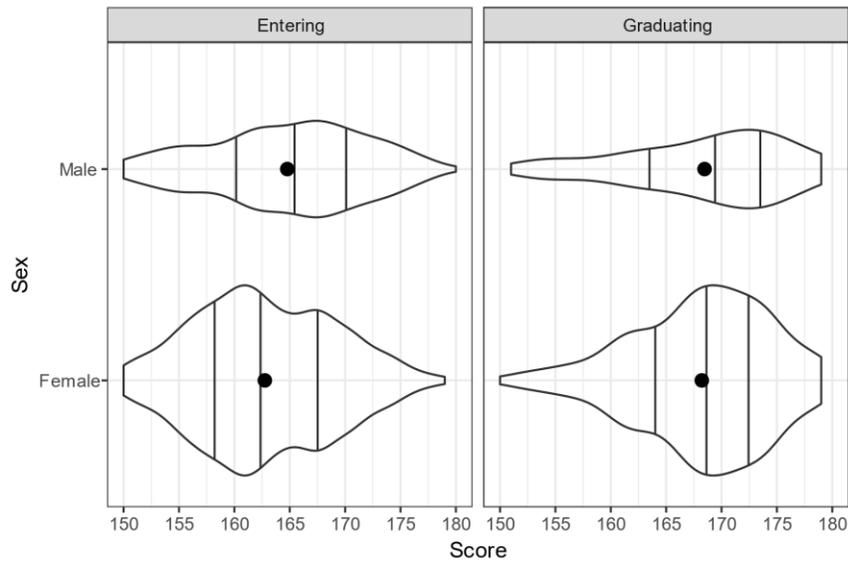
Test Scores by Selected Characteristics

Turning our attention to differences by sex, we note that females constitute the majority of the sample for both the entering and the graduating class. Males tend to do better on the test for both the entering and graduating class, especially for the entering class.

The bottom quartile is slightly lower for males in the graduating class. The mean scores for males and females are almost identical, but the median is about two points higher for males than females in the graduating class.

The general pattern in HEIghten scores by class repeats when the results are broken down by sex, with both males and females generally scoring better in the graduating class than the entering class.

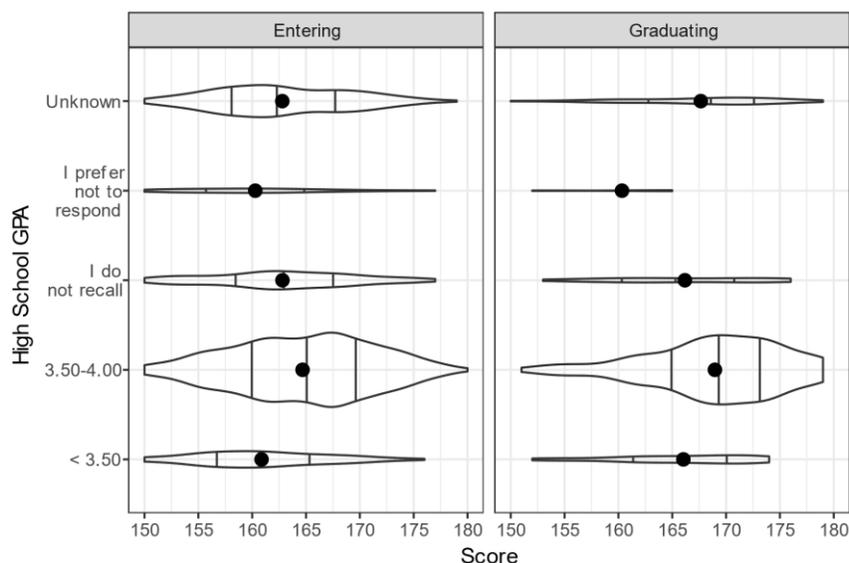
Figure 17: Violin Plots of HEIghten Scores by Cohort and Sex, University



Not surprisingly, there is a positive association between high school GPA and test scores: Entering students with a GPA below 3.5 had a mean score roughly 4 points below those with a high school GPA of 3.5–4. Similarly, median, upper and lower quartile scores were both lower, as well. Students who chose not to respond to this question, could not recall their GPA, or for whom this information was missing also did not do as well and the missing category was in fact the second most common group (after those with a GPA of 3.5–4) for the entering class.

Students who reported that they attained a GPA below 3.5 were therefore rare in the entering class and these students constituted a relatively smaller part of the graduating sample. As for the entering class, among students who reported their high school GPA, those with lower reported GPAs performed worse on HEIghten. Students who chose not to respond, could not recall, or had an otherwise unknown high school GPA were also less common in the graduating class. Nonetheless, students in these categories also had lower scores than those who reported that they attained a GPA between 3.5 and 4 across the range of the summary statistics shown.

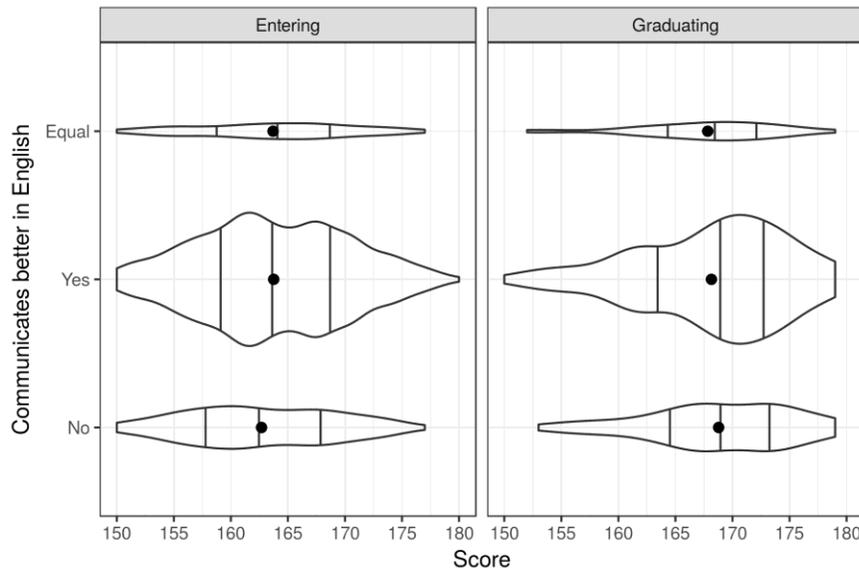
Figure 18: Violin Plots of HEIghten Scores by Cohort and High School GPA, University



Students who responded that they communicate better in English than another language scored only slightly better on the test than those who felt they were more comfortable in another language in the entering class. The summary statistics for those who reported they communicated better in English and those who evaluated their skill to be equal in English and another language were similar.

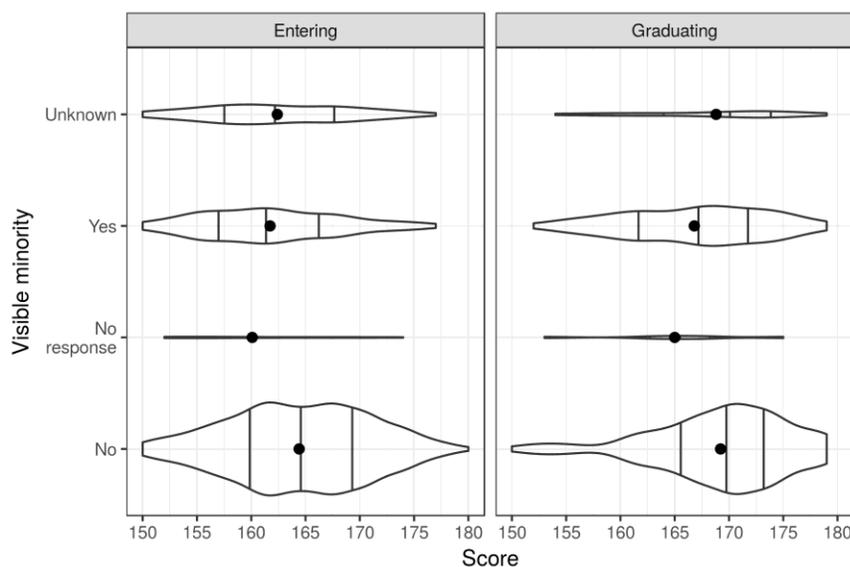
Among the graduating class, students who felt most comfortable with English had slightly lower quartile than other students. The mean scores for the three groups are more or less similar, with those that note they are *not* as comfortable in English as another language having only a slightly higher mean and upper quartile score than others.

Figure 19: Violin Plots of HEIghten Scores by Cohort and Language, University



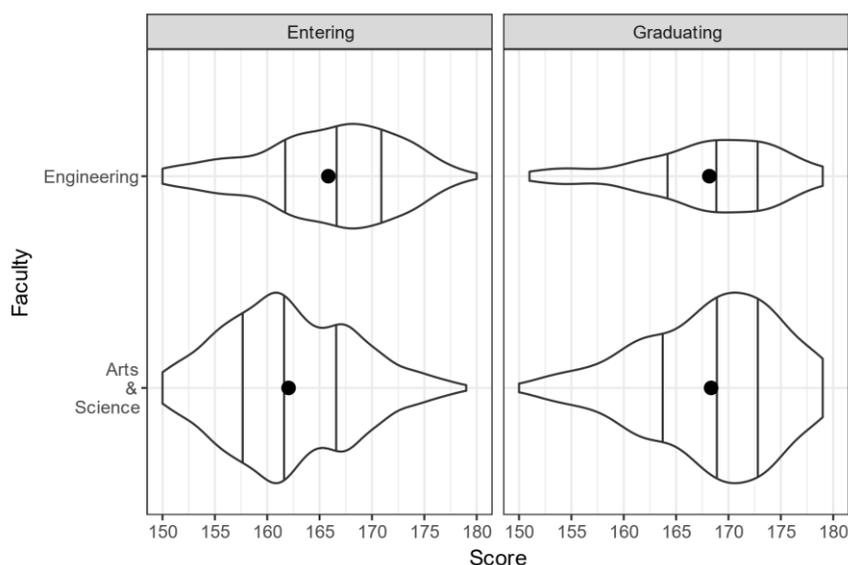
Looking at the results based on whether a student is a visible minority or not, we note that the majority of the students reported themselves as not being a visible minority, followed by those who did report themselves as a visible minority. There were very few students who did not respond to this question both in the entering and graduating class and unknown responses were also relatively rare, especially for the graduating class. Once again, all of the groups did better in the graduating than entering class. Visible minorities appear to score lower than other students among both the entering and graduating students.

Figure 20: Violin Plots of HEIghten Scores by Cohort and Visible Minority Status, University



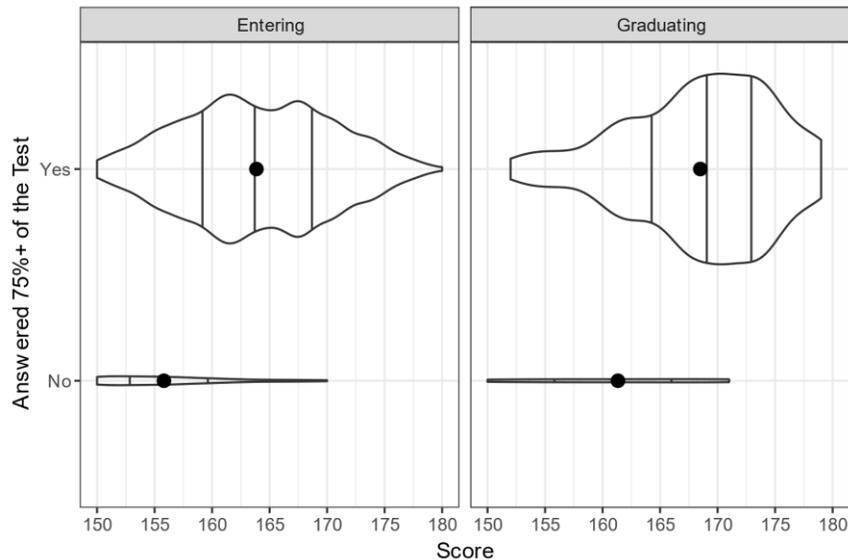
Due to the nature of the entering year at the university and the fact that students do not declare their major at the outset of their studies, we could only compare the performance of the Faculty of Arts and Science and the Faculty of Engineering, without assessing program-level differences. Nonetheless, the results are instructive. Graduating students once again score higher than entering students on HEIghten in both faculties. The two faculties appear to have similar summary statistics for the graduating students, with only slightly higher median and lower upper and lower quartiles for Arts and Science. This is not the case with the entering class, where the scores for the Faculty of Arts and Science appear to be lower than those of the Faculty of Engineering across all the summary measures.

Figure 21: Violin Plots of HEIghten Scores by Cohort and Faculty, University



As we would expect, those who answered less than 75% of the test questions appear to score lower compared to those who answered more questions for both the entering and graduating class. However, the vast majority of students in both classes completed more than 75% of the test. Once again, the graduating students do better than the entering students; even among the small proportion of students that answered less than 75% of the questions, the mean score is over 5 points higher in the graduating group. This is roughly the same score advantage we observe in the overall comparison in mean test scores between the two groups of students in each year.

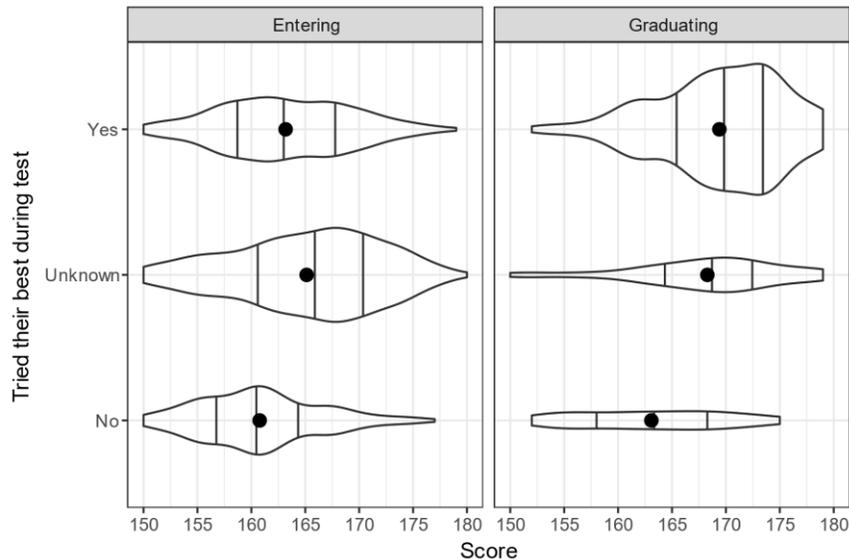
Figure 22: Violin Plots of HEIghten Scores by Cohort and Test Completion, University



Students were also surveyed as to whether they tried their best during the test. The proportion of students who answered “no” for this question appears to be relatively larger in the entering class than the graduating class. Interestingly, the proportion of students who did not answer is relatively larger in the entering year, whereas the unknown category is relatively uncommon for the graduating class.

Those who tried their best actually score *worse* than the unknown category among the entering class. This is not the case for the graduating class, where the two groups have more or less similar performance. Furthermore, the difference in mean scores between those who say they did not try their best and those who say they did appears to be larger for the graduating students than the entering students. Most of the graduating class reported they tried their best during the test.

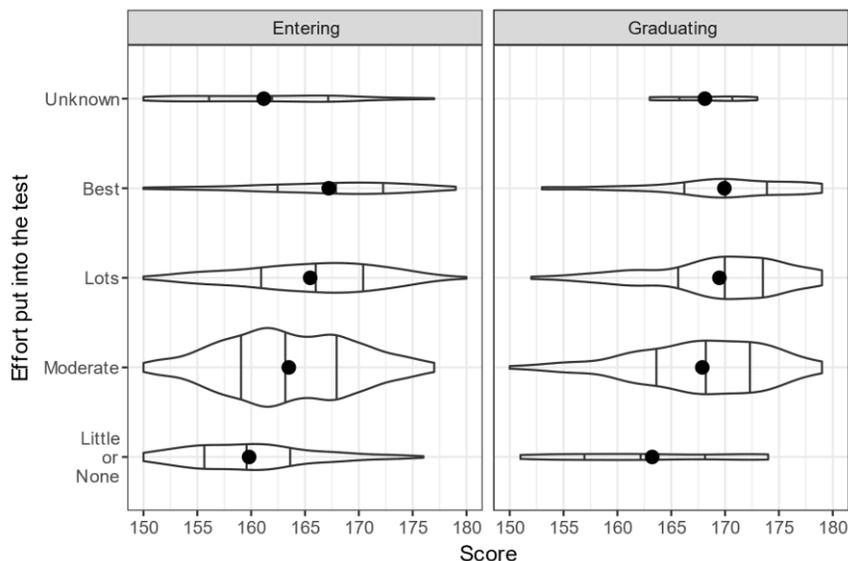
Figure 23: Violin Plots of HEIghten Scores by Cohort and Whether Students Tried Their Best, University



Students also gauged the amount of effort they put into the test. In both classes, most students reported as having put at least a moderate effort into the test. While students who put little or no effort were not common among the graduating students, they were relatively more common among the entering students. Students who reported they put their best effort into the test constituted a relatively small part of the sample for both the entering and graduating class. As elsewhere, students in the graduating class did better on HEIghten for the entire range of categories possible when compared to the entering class.

Among those who answered this question, there seems to be a positive association between the level of effort as reported by students and the HEIghten score for both the entering and graduating students.

Figure 24: Violin Plots of HEIghten Scores by Cohort and Effort, University



5.3 Regression Analysis

Entering Cohort

We first look at Table R4, which shows a series of regressions where the dependent variable is the HEIghten test scores for the entering class of university students. We again run three separate models for this group: a characteristics model, which in this case includes sex, high school GPA, whether the student communicates better in English or another language (or equally well in both languages), and whether the student is a visible minority. In addition, we run a second regression that includes all of the explanatory variables from the characteristics regression but also includes the student’s faculty. Finally, the third model (“Effort”) adds whether the student answered 75% or more of the test questions and also the student’s answers to the effort-based questions included in the HEIghten instrument, namely whether the student tried their best or not during the test and also the amount of effort the student put into the test on a scale from “little or none” to “best.”

Table 2, seen earlier, where HEIghten scores are reported on a simple variable-by-variable basis, shows a gender gap in mean scores of about 2 points. This is close to what we see in the first column in Table R4, where males on average score about 1.9 points higher on the HEIghten test than females when all other student characteristics are taken into account. This suggests that the gender gap in scores is not explained by differences in high school GPA, language or visible minority status. However, the gender gap in scores is greatly reduced and is no longer statistically significant when the faculty variable (i.e., Engineering) is added (model 2), indicating that the gender differences are related to differences in the faculties in which males and females are enrolled (i.e., men are more likely to be in the high-scoring Engineering faculty than women). Adding effort has little additional effect on the estimated difference. In summary, and as

previously seen for the college students, the male-female difference in HEIghten scores appears to be driven principally by differences in the programs/faculties in which males and females enrol.⁷

Those who attained a GPA between 3.5 and 4 in high school (i.e., the highest level) score more than 3 points higher than those with a GPA lower than 3.5 (the omitted/comparison group). Students whose high school GPA is not known score 2.2 points higher and those who do not recall score 1.3 points higher, while the results for those who preferred not to respond are very small and not statistically significant. The reported difference in scores between those who report having a GPA within the range 3.5–4 and those who report a lower GPA in high school is somewhat reduced but largely persists even when faculty and effort variables are included to the model, suggesting that the association between high school GPA and HEIghten score cannot be explained by the characteristics in the data and represents a general effect. (The effect for those whose GPA is unknown is essentially unchanged across the different models.)

Students who report communicating better in English, or equally well in English and another language are estimated to score a little higher (between half a point and just over 1 point) than those who report communicating better in another language across all models. However, the differences in scores between these groups are not statistically significant in any of the models, except for the “equal” group.

Students who identify as a visible minority score around 2.5 points lower, on average, than non-minorities and the effect is little changed across models. Those who do not respond or for whom minority status is unknown also generally score lower than non-minorities.

In model 2, Engineering students are seen to score about 3.4 points higher, on average, than those in the Faculty of Arts and Science. This difference decreases to about 2 points when effort is added to the model, suggesting that part of the faculty difference in the HEIghten scores is explained by the fact that Engineering students put more effort into the test than Arts and Science students.⁸

Unsurprisingly, the number of answered questions is highly important, with those who completed at least 75% of the test scoring 7.7 points higher than those who did not.

There is very little difference in scores among those who say they tried their best, those for whom the answer is unknown and those who report not trying their best and none of the estimates are statistically significant.

Self-reported effort (a different variable) appears to be a strong indicator of how well a student scores.⁹ Students who reported putting moderate effort into the test score about 2.7 points higher than those who report putting little or no effort into the test. Those who report putting “lots” of effort score about 4.5 points higher and those who report putting their best effort score almost 6 points higher than those who report putting little or no effort.

7 In the case of college students, the gender difference is also related to differences in assessment scores.

8 Around 40% of the Engineering students report either “lots” or “best” for effort put into the test, whereas this percentage is around 22% for Arts and Science students.

9 The R² value increases from 0.13 to 0.25 when the effort-related variables are included in the model.

Overall, the results suggest that for the entering university students, scores appear to be related to high school GPA, communicating better in English, number of questions answered on the test and the degree of effort put into the test. Furthermore, the gender gap in scores appears to be mainly accounted for by the fact that males and females enrol in programs in different faculties (i.e., more males in Engineering), while most of the other patterns in scores are unaffected when faculty and effort variables are included in the model.

Graduating Cohort

For the graduating class of university students, we report results for the same set of regressions as for the entering class: characteristics, faculty and effort (Table R5).

Unlike what we see with our entering class, when student characteristics are taken into account (the first model), males do not score higher than females; the coefficient is positive (by three-quarters of a point) but is not statistically significant. In contrast, including effort in the model increases the gender gap in scores to 2 points, indicating that males tend to put less effort into the test than females: that is, controlling for that low effort, they do better.¹⁰

The high school grade advantage remains, with students who attained a high school GPA of 3.5 to 4 scoring about 3 points higher on the test than those with a GPA below 3.5. This result holds when faculty and effort variables are added to the model. These results are consistent with the entering cohort and point to GPA having a general effect.

Furthermore, similar to the positive relationship between HEIghten scores and the reading, writing and math assessments found in the college analysis, the GPA results represent a kind of validation of the HEIghten tool: While GPA is not a measure of critical thinking *per se*, one would expect HEIghten scores to be positively related to GPA, which captures students' ability in a different way.

Interestingly, whether a student communicates better in English or another language no longer seems to matter for the graduating class, unlike the entering class. By the later point in their studies, students' language skills may have improved such that they are no longer related to their HEIghten scores. This would support the notion that for the entering class, language is possibly a marker of having problems with taking the test rather than capturing critical thinking as such.

Conversely, students who identify as a visible minority score a statistically significant 2.7 points lower, on average, than those who do not identify as visible minority — a result that holds in the other models (faculty and effort).

There are no statistically significant differences in scores by faculty.

¹⁰ Around 19% of males and 37% of females say they put lots of effort into the test, and around 60% of males and 71% of females say they tried their best on the test.

Turning to the variables introduced in model 4 (“Effort”), we see that those who finished 75% or more of the test score almost 8 points higher than those who answered fewer questions. Similarly — and unlike the entering students — those who say they tried their best or for whom the answer to this question is unknown score considerably higher (around 5 points) than those who did not try their best. Finally, the degree of effort put into the test (a separate question) is also strongly related to HEIghten scores.

The importance of these latter three sets of measures — whether 75% or more of the test was completed and the two effort questions — the R^2 value of the model (representing how much of the variance in outcomes is explained by the variables included in the model) increases from 0.08 to 0.25.

In summary, the results suggest that for the graduating university students, scores appear to be related to gender, high school GPA, visible minority status, number of questions answered on the test, whether the students tried their best and the degree of effort put into the test. Unlike the entering students, the gender gap in scores persists in the final model, which includes all variables, suggesting that the male-female difference in scores cannot be explained by differences in the other variables included in the analysis. Neither faculty nor the effort-related variables generally change the patterns in scores with respect to the other variables included in the analysis characteristics.

Lastly, as also seen for the college students, there appears to be some differences in how HEIghten scores are related to various student and program characteristics and effort, which implies that skill level (or HEIghten score) differences between the classes may differ across certain groups of students. This will be examined and tested further in the next subsection using a model fully interacted with the class variable (Table R6).

Cross-cohort Comparisons

Table R6 shows results from a pooled regression model, which includes both the entering and graduating students comparable to those included in the college analysis. Again, all of the explanatory variables included in the previous section remain, but we also include sets of “class” explanatory variables that differentiate whether a student belongs to the entering or graduating class in order to identify how the differences in HEIghten scores across the two classes differ across the variables included in the analysis.

The first model (class) simply shows that the graduating class scores 4.8 points higher, on average, than the entering class when no other explanatory variables appear in the model, a result that mirrors the descriptive statistics.

The other models again show how the differences in scores between the graduating and entering classes differ across the variables included in the models. For example, the coefficient estimate on the “Graduating * Engineering” variable is -3.84 in the third model (Faculty), which means that the difference between the two classes is smaller for engineers than for students in Arts and Science, holding all other factors constant. This difference in the skill difference between classes diminishes in magnitude (to 1.84 points) when effort variables are included in the regression.

The other most significant difference in how skills differ across classes is seen in the strong coefficient on the “tried your best” interaction. In particular, those who tried their best showed a larger difference (by 4.83 points) in scores between the two classes compared to those who did not try their best.

6. Conclusion

This paper presents the results of an analysis of students’ HEIghten critical-thinking scores at a large college and a large university, both located in Ontario. The goals of this study are to:

1. Measure and compare entering and graduating students’ critical-thinking levels; and
2. Identify the relationships between critical-thinking scores and a range of student and program characteristics.

Using descriptive and regression approaches, the main results of the analysis are as follows:

College results:

- There is no statistically significant difference in the HEIghten scores between the graduating and entering class.
- Students who are male, whose first language is English or French, who are not visible minorities, who did well on entry assessment tests (reading, writing and math) and who spent more time on the test tend to record higher critical-thinking scores in both classes.
- Students age 24–26 at entry tend to score higher than those age 18 and under in the entering class, and students age 22–23 at entry tend to score higher than those age 18 and under in the graduating class.
- Scores appear to vary across programs. These differences are relatively small for the entering class, but are larger for the graduating class. Community and justice students tend to score lower than students in other programs, the only exception being the scores of business students in the entering class.
- Based on a regression analysis, these patterns still hold when differences in student and program characteristics, assessment scores and time spent on the test are taken into account, but the magnitudes of score differences by the various characteristics decrease, with the greatest effect coming from the assessment scores.

- Based on a regression analysis, which combines the entering and graduating students, differences in skill levels between the graduating and entering classes vary across age and programs, with students age 22–23 showing a larger difference in skills between the graduating and entering classes than other age groups and community and justice studies students showing a smaller difference between graduating and entering classes than other programs. The differences in scores between classes are larger for those who scored within the second and sixth category of the reading assessment scores, smaller for those who scored in the second and third quartiles of the writing assessment score distribution, and larger for those who scored in the second and third quartiles of the math assessment score distribution compared to those who scored in the bottom category or quartile. No other differences between the graduating and entering classes across the range of student characteristics, program and the other variables included in the analysis are statistically significant.

University results:

- Overall, graduating students achieve critical-thinking scores 4.8 points higher than the entering students, 168.3 and 163.5 respectively.
- For the entering class, students who are male, who have higher high school GPAs, who communicate better in English, who are not a visible minority, who are studying engineering and who put more effort into the test score higher on average.
- For the graduating class, students who have higher high school GPAs, who are not a visible minority and who put more effort into the test record higher scores.
- The regression analysis shows that the score patterns generally do not change when all other factors (student characteristics, faculty and effort) are taken into account, except for gender in the graduating class: Taking into account other student characteristics, faculty and the effort put into the test, male students tend to score higher than their female counterparts.
- The gap in critical-thinking scores between the graduating and entering classes is smaller for Faculty of Engineering students than Faculty of Arts and Science students and larger for students who tried their best than for students who did not. No other differences between the classes across the student characteristics and the other variables are statistically significant.

We caution, however, that comparisons of the entering and graduating students included in this analysis (which represent two different sets of students assessed at a single point in time) may be driven by various factors. For instance, critical-thinking scores may be influenced by differences in the characteristics of the students in the entering year and those who made it to their graduating year, by differences in the samples of students who took the test in the two years stemming from how students were recruited for the tests at each institution, by any underlying differences in the students in the two classes at each institution and potentially by other factors as well, as discussed at length in the report.

Therefore, we suggest an alternative approach that could be adopted for future studies of skill levels, changes in skill levels during PSE and how these changes may be related to a range of student characteristics and schooling experiences.

In particular, we suggest, where feasible, following students through their studies and administering assessments at different points in time so that skill levels — in particular changes in skill levels — could be identified for given sets of students, as opposed to comparing skill levels across different sets of students who may differ in a range of ways related to skill levels as discussed above.

Using a longitudinal approach would not only allow skill levels to be measured at any point in time and skill gains to be identified as students moved through PSE, but would also allow skill levels and skill gains to be related to student characteristics, the programs they take and, potentially, their specific schooling experiences, thus addressing the key question mentioned above regarding, “Which of these skills can be developed or improved in PSE, and if they can, how?”

We also suggest relating skill levels to student outcomes while in PSE, such as grades and progress to graduation, thus providing a fuller perspective of the impacts and value of skills and this would be more meaningful if skill levels were measured at different points in time so that subsequent PSE outcomes could be linked to those measures.

Such a longitudinal approach would require considerable resources to implement and obtaining results would take time as students were followed through their studies, ultimately from beginning to end — although “synthetic” profiles of skill levels and year-to-year skill gain could be obtained by following students in each year to the next, thus identifying skill levels in any given year and skill changes from each year to the next, for different sets of students (e.g., first-year students could be followed into second year, second-year students into third year and so on).

Nevertheless, the PAWS project begins to fill an important gap in our understanding of skills in the PSE context by measuring skill levels of entering and graduating students, while also developing an initial analytical framework that could be useful for future research of this type.

References

- Barrington, L. & Casner-Lotto, J. (2006). Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st century U.S. workforce. United States: The Conference Board.
- Business Council of British Columbia. (2010). 2010 Biennial skills and attributes survey report: What are BC employers looking for? Vancouver: Business Council of British Columbia.
- Business Council of Canada. (2016), Developing Canada's future workforce: A survey of large private-sector employers.
- Canadian Association of Career Educators and Employers, 2013. Campus Recruitment Report.
- Gomathy, S. 2017. Developing employability skills among the youth of present generation: Soft skills to enhance English speaking skill. SSRG *International Journal of Humanities and Social Science (ICEHS), Special Issue*, May 2017, 1–8
- Johnson, P. (2009). The 21st Century Skills Movement. *Educational Leadership*, 67(1), 11.
- Liu, L., Mao, L., Frankel, L. & Xu, J. (2016). Assessing critical thinking in higher education: The HEIghten™ approach and preliminary validity evidence. *Assessment & Evaluation in Higher Education*. 41(5), 677–694.
- Royal Bank of Canada, 2018. Humans Wanted.

Regression Results 1: College

Table R1: Test Score Regression Results, Entering Students, College

	Characteristics	Program	Assessments	Time Spent
(Intercept)	158.30*** (0.39)	157.82*** (0.51)	153.27*** (0.6)	149.82*** (0.69)
<i>Sex (Female)</i>				
Male	1.69*** (0.37)	1.22*** (0.39)	0.69** (0.34)	0.64** (0.32)
<i>Age at Entry (1-18)</i>				
19	0.7 (0.46)	0.82* (0.46)	0.53 (0.38)	0.36 (0.37)
20-21	1.86*** (0.53)	2.01*** (0.53)	0.71 (0.44)	0.6 (0.41)
22-23	1.69*** (0.69)	1.73*** (0.69)	0.08 (0.6)	0.23 (0.56)
24-26	3.76*** (0.84)	3.82*** (0.84)	1.42** (0.71)	1.44** (0.67)
27+	2.57*** (0.69)	2.54*** (0.69)	0.18 (0.58)	-0.63 (0.57)
<i>Mother Tongue (Official First Language)</i>				
Other First Language	-3.20*** (0.64)	-3.24*** (0.64)	-1.24* (0.65)	-1.58*** (0.63)
<i>First Generation PSE (No)</i>				
Yes	-0.97** (0.41)	-0.96** (0.41)	0.06 (0.34)	0.09 (0.32)
<i>Visible minority (No)</i>				
No response	-2.27*** (0.92)	-2.39*** (0.92)	-0.74 (0.74)	-0.59 (0.7)
Yes	-1.98*** (0.42)	-2.09*** (0.42)	-1.11*** (0.35)	-1.26*** (0.34)
Unknown	0.35 (0.59)	0.34 (0.59)	0.25 (0.5)	0.06 (0.46)
<i>Program (Business)</i>				
Community and Justice		0.3 (0.47)	0.5 (0.48)	0.56 (0.47)

Engineering	1.66*** (0.51)	1.03** (0.43)	0.79** (0.41)
Media and Entertainment	0.76 (0.67)	-0.09 (0.6)	-0.00 (0.59)
<hr/>			
<i>Reading Assessment (1)</i>			
2		0.24 (0.5)	-0.08 (0.5)
3		1.93*** (0.54)	1.41*** (0.53)
4		2.33*** (0.53)	1.45*** (0.52)
5		4.10*** (0.53)	3.21*** (0.53)
6		4.27*** (0.6)	3.52*** (0.58)
7		7.12*** (0.6)	6.33*** (0.59)
8		9.46*** (0.7)	8.59*** (0.68)
(Missing)		4.86*** (0.87)	2.96*** (0.74)
<hr/>			
<i>Writing Assessment (1)</i>			
2		0.59* (0.35)	0.70** (0.34)
3		1.75*** (0.47)	1.68*** (0.46)
4		1.38* (0.79)	1.74** (0.74)
(Missing)		-3.23*** (0.84)	-2.13*** (0.68)
<hr/>			
<i>Math Assessment (1)</i>			
2		-0.32 (0.58)	-0.42 (0.55)
3		0.84 (0.61)	0.58 (0.58)
4		2.19*** (0.68)	1.66*** (0.65)
(Missing)		0.88 (0.57)	-0.01 (0.55)

Minutes Spent				0.14*** (0.01)
N	1071	1071	1071	1071
R2	0.1	0.11	0.4	0.45

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table R2: Test Score Regression Results, Graduating Students, College

	Characteristics	Program	Assessments	Time Spent
(Intercept)	159.87*** (0.68)	160.14*** (0.76)	153.66*** (1.05)	149.08*** (1.21)
<i>Sex (Female)</i>				
Male	1.55*** (0.51)	1.01** (0.52)	0.76* (0.45)	0.78* (0.42)
<i>Age at Entry (1-18)</i>				
19	-0.46 (0.81)	-0.37 (0.78)	-0.22 (0.68)	-0.29 (0.64)
20-21	1.24 (0.77)	1.41* (0.76)	1.21* (0.65)	1.29** (0.6)
22-23	3.70*** (0.97)	3.60*** (0.95)	2.32*** (0.85)	2.07*** (0.82)
24-26	0.45 (0.94)	0.67 (0.96)	0.43 (0.87)	-0.05 (0.83)
27+	1.99*** (0.81)	2.15*** (0.78)	0.95 (0.71)	-0.15 (0.72)
<i>Mother Tongue (Official First Language)</i>				
Other First Language	-2.48*** (0.6)	-2.50*** (0.6)	-1.79*** (0.64)	-1.42** (0.62)
<i>First Generation PSE (No)</i>				
Yes	-0.95* (0.58)	-0.56 (0.57)	-0.09 (0.49)	0.08 (0.48)
<i>Visible minority (No)</i>				
No response	-3.79*** (1.15)	-3.91*** (1.22)	-1.79 (1.18)	-2.40** (1.23)
Yes	-2.51*** (0.72)	-2.70*** (0.69)	-1.15* (0.64)	-1.75*** (0.62)
Unknown	-1.92*** (0.68)	-2.22*** (0.66)	-1.17** (0.59)	-1.44*** (0.57)
<i>Program (Business)</i>				
Community and Justice		-1.55*** (0.59)	-1.76*** (0.58)	-1.72*** (0.56)
Engineering		1.08* (0.63)	0.41 (0.59)	0.51 (0.56)
Media and Entertainment		2.95** (1.48)	1.86 (1.20)	1.64 (1.23)

<i>Reading Assessment (1)</i>				
2			2.23**	1.93*
			(1.13)	(1.08)
3			3.60***	3.17***
			(1.07)	(1.07)
4			3.24***	3.00***
			(0.88)	(0.86)
5			3.79***	3.94***
			(1.00)	(0.98)
6			6.50***	6.26***
			(1.1)	(1.07)
7			6.54***	6.18***
			(0.98)	(1.02)
8			10.80***	10.33***
			(1.11)	(1.12)
(Missing)			0.05	-0.15
			(1.45)	(1.52)
<i>Writing Assessment (1)</i>				
2			-0.86	-1.11
			(0.76)	(0.73)
3			-0.05	-0.39
			(0.78)	(0.74)
4			0.25	0.40
			(1.17)	(1.12)
(Missing)			2.96**	2.82*
			(1.45)	(1.48)
<i>Math Assessment (1)</i>				
2			2.88***	2.50***
			(0.98)	(0.91)
3			3.56***	3.22***
			(0.82)	(0.76)
4			1.46*	1.51*
			(0.88)	(0.83)
(Missing)			1.93***	1.96***
			(0.72)	(0.68)
Minutes Spent				0.14***
				(0.02)
N	484	484	484	484
R2	0.17	0.21	0.40	0.45

*** p < 0.01; ** p < 0.05; * p < 0.1.

Table R3: Test Score Regression Results, Combined Classes with Interactions, College

	Class	Characteristics	Program	Assessments	Time Spent
(Intercept)	159.30*** (0.19)	158.30*** (0.39)	157.82*** (0.51)	153.27*** (0.60)	149.82*** (0.69)
<i>Class (Entering)</i>					
Graduating	-0.14 (0.33)	1.57** (0.78)	2.32*** (0.92)	0.39 (1.21)	-0.74 (1.4)
<i>Sex (Female)</i>					
Male		1.69*** (0.37)	1.22*** (0.39)	0.69** (0.34)	0.64** (0.32)
<i>Age at Entry (1-18)</i>					
19		0.70 (0.46)	0.82* (0.46)	0.53 (0.38)	0.36 (0.37)
20-21		1.86*** (0.53)	2.01*** (0.53)	0.71 (0.44)	0.6 (0.41)
22-23		1.69*** (0.69)	1.73*** (0.69)	0.08 (0.6)	0.23 (0.56)
24-26		3.76*** (0.84)	3.82*** (0.84)	1.42** (0.71)	1.44** (0.67)
27+		2.57*** (0.69)	2.54*** (0.69)	0.18 (0.58)	-0.63 (0.57)
<i>Mother Tongue (Official First Language)</i>					
Other First Language		-3.20*** (0.64)	-3.24*** (0.64)	-1.24* (0.65)	-1.58*** (0.63)
<i>First Generation PSE (No)</i>					
Yes		-0.97** (0.41)	-0.96** (0.41)	0.06 (0.34)	0.09 (0.32)
<i>Visible minority (No)</i>					
No response		-2.27*** (0.92)	-2.39*** (0.92)	-0.74 (0.74)	-0.59 (0.70)
Yes		-1.98*** (0.42)	-2.09*** (0.42)	-1.11*** (0.35)	-1.26*** (0.34)
Unknown		0.35 (0.59)	0.34 (0.59)	0.25 (0.5)	0.06 (0.46)
<i>Class (Entering), Sex (Female)</i>					
Graduating * Male		-0.13 (0.63)	-0.21 (0.65)	0.08 (0.56)	0.14 (0.53)

<i>Class (Entering), Age at Entry (1-18)</i>				
Graduating * 19	-1.15 (0.94)	-1.19 (0.91)	-0.76 (0.78)	-0.65 (0.74)
Graduating * 20-21	-0.62 (0.94)	-0.60 (0.92)	0.5 (0.78)	0.68 (0.73)
Graduating * 22-23	2.01* (1.19)	1.87 (1.18)	2.24** (1.04)	1.84* (0.99)
Graduating * 24-26	-3.31*** (1.26)	-3.15*** (1.28)	-0.99 (1.12)	-1.49 (1.06)
Graduating * 27+	-0.58 (1.07)	-0.39 (1.04)	0.77 (0.91)	0.48 (0.92)
<i>Class (Entering), Mother Tongue (Official First Language)</i>				
Graduating * Other First Language	0.72 (0.87)	0.74 (0.88)	-0.55 (0.91)	0.16 (0.89)
<i>Class (Entering), First Generation PSE (No)</i>				
Graduating * Yes	0.02 (0.71)	0.4 (0.7)	-0.15 (0.59)	-0.01 (0.58)
<i>Class (Entering), Visible minority (No)</i>				
Graduating * No response	-1.52 (1.48)	-1.52 (1.53)	-1.05 (1.39)	-1.82 (1.41)
Graduating * Yes	-0.53 (0.83)	-0.60 (0.81)	-0.04 (0.73)	-0.49 (0.71)
Graduating * Unknown	-2.27*** (0.90)	-2.56*** (0.89)	-1.42* (0.77)	-1.51** (0.73)
<i>Program (Business)</i>				
Community and Justice		0.3 (0.47)	0.5 (0.48)	0.56 (0.47)
Engineering		1.66*** (0.51)	1.03** (0.43)	0.79** (0.41)
Media and Entertainment		0.76 (0.67)	-0.09 (0.6)	-0.00 (0.59)
<i>Class (Entering), Program (Business)</i>				
Graduating * Community and Justice		-1.85*** (0.76)	-2.26*** (0.75)	-2.28*** (0.73)
Graduating * Engineering		-0.58 (0.81)	-0.62 (-0.73)	-0.28 (0.69)

Graduating * Media and Entertainment	2.19 (1.62)	1.94 (1.34)	1.64 (1.36)
<hr/>			
<i>Reading Assessment (1)</i>			
2		0.24 (0.50)	-0.08 (0.50)
3		1.93*** (0.54)	1.41*** (0.53)
4		2.33*** (0.53)	1.45*** (0.52)
5		4.10*** (0.53)	3.21*** (0.53)
6		4.27*** (0.60)	3.52*** (0.58)
7		7.12*** (0.60)	6.33*** (0.59)
8		9.46*** (0.70)	8.59*** (0.68)
(Missing)		4.86*** (0.87)	2.96*** (0.74)
<hr/>			
<i>Writing Assessment (1)</i>			
2		0.59* (0.35)	0.70** (0.34)
3		1.75*** (0.47)	1.68*** (0.46)
4		1.38* (0.79)	1.74** (0.74)
(Missing)		-3.23*** (0.84)	-2.13*** (0.68)
<hr/>			
<i>Math Assessment (1)</i>			
2		-0.32 (0.58)	-0.42 (0.55)
3		0.84 (0.61)	0.58 (0.58)
4		2.19*** (0.68)	1.66*** (0.65)
(Missing)		0.88 (0.57)	-0.01 (0.55)
<hr/>			

<i>Class (Entering), Reading Assessment (1)</i>		
Graduating * 2	1.99 (1.24)	2.00* (1.19)
Graduating * 3	1.67 (1.20)	1.76 (1.19)
Graduating * 4	0.9 (1.02)	1.54 (1.01)
Graduating * 5	-0.31 (1.13)	0.73 (1.12)
Graduating * 6	2.23* (1.25)	2.74** (1.22)
Graduating * 7	-0.58 (1.15)	-0.14 (1.18)
Graduating * 8	1.34 (1.31)	1.73 (1.31)
Graduating * (Missing)	-4.81*** (1.69)	-3.11* (1.69)
<i>Class (Entering), Writing Assessment (1)</i>		
Graduating * 2	-1.45* (0.84)	-1.81** (0.81)
Graduating * 3	-1.79** (0.91)	-2.07** (0.87)
Graduating * 4	-1.13 (1.41)	-1.34 (1.34)
Graduating * (Missing)	6.18*** (1.68)	4.95*** (1.63)
<i>Class (Entering), Math Assessment (1)</i>		
Graduating * 2	3.20*** (1.14)	2.92*** (1.06)
Graduating * 3	2.72*** (1.02)	2.64*** (0.96)
Graduating * 4	-0.73 (1.11)	-0.15 (1.05)
Graduating * (Missing)	1.06 (0.92)	1.97** (0.88)
Minutes Spent		0.14*** (0.01)

<i>Class (Entering), Minutes Spent</i>					
Graduating * Minutes Spent					0.01 (0.03)
N	1555	1555	1555	1555	1555
R2	0	0.12	0.14	0.4	0.45

*** p < 0.01; ** p < 0.05; * p < 0.1.

Regression Results 2: University

Table R4: Test Score Regression Results, Entering Students, University

	Characteristics	Faculty	Effort
(Intercept)	160.94*** (0.64)	160.64*** (0.63)	150.17*** (1.14)
<i>Sex (Female)</i>			
Male	1.88*** (0.42)	0.54 (0.44)	0.66 (0.41)
<i>High School GPA (< 3.50)</i>			
3.50-4.00	3.12*** (0.64)	2.20*** (0.64)	2.09*** (0.59)
I do not recall	1.30* (0.8)	0.88 (0.78)	1.07 (0.69)
I prefer not to respond	-0.41 (1.19)	-0.80 (1.14)	-1.34 (1.14)
Unknown	2.23*** (0.75)	2.19*** (0.74)	2.34*** (0.69)
<i>Communicates better in English (No)</i>			
Yes	0.48 (0.44)	0.52 (0.43)	0.55 (0.41)
Equal	1.33* (0.77)	1.30* (0.74)	1.33** (0.69)
<i>Visible minority (No)</i>			
No response	-3.66 (2.36)	-5.12** (2.38)	-2.47 (2.69)
Yes	-2.54*** (0.47)	-2.52*** (0.47)	-2.29*** (0.45)
Unknown	-1.65** (0.70)	-1.34** (0.69)	-1.36** (0.65)
<i>Faculty (Arts & Science)</i>			
Engineering		3.36*** (0.46)	2.04*** (0.82)
<i>Answered 75%+ of the Test (No)</i>			
Yes			7.73*** (0.93)
<i>Tried their best during test (No)</i>			
Unknown			0.8 (0.79)
Yes			0.59 (0.53)
<i>Effort put into the test (Little or None)</i>			
Moderate			2.70*** (0.54)
Lots			4.46*** (0.66)

Students Measuring Critical-thinking Skills of Postsecondary Students

Best			5.96*** (0.87)
Unknown			3.01*** (0.97)
N	1075	1075	1075
R2	0.09	0.13	0.25

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table R5: Test Score Regression Results, Graduating Students, University

	Characteristics	Faculty	Effort
(Intercept)	167.00*** (1.21)	167.07*** (1.21)	150.96*** (2.97)
<i>Sex (Female)</i>	0.75 (0.81)	0.93 (0.82)	1.99*** (0.71)
Male			
<i>High School GPA (< 3.50)</i>			
3.50-4.00	3.04*** (1.15)	3.06*** (1.15)	3.09*** (1.13)
I do not recall	0.45 (1.76)	0.53 (1.76)	0.73 (1.64)
I prefer not to respond	-5.71 (3.95)	-5.79 (3.96)	-5.63 (4.22)
Unknown	1.32 (2.18)	1.33 (2.18)	1.68 (1.92)
<i>Communicates better in English (No)</i>			
Yes	-0.55 (0.79)	-0.55 (0.78)	-0.65 (0.75)
Equal	0.42 (1.38)	0.35 (1.39)	0.51 (1.34)
<i>Visible minority (No)</i>			
No response	-3.28 (3.17)	-3.14 (3.17)	-2.43 (2.58)
Yes	-2.68*** (0.82)	-2.70*** (0.81)	-2.63*** (0.71)
Unknown	0.15 (2.56)	0.15 (2.57)	0.65 (2.10)
<i>Faculty (Arts & Science)</i>			
Engineering		-0.47 (0.80)	0.2 (0.74)
<i>Answered 75%+ of the Test (No)</i>			
Yes			7.91*** (2.25)
<i>Tried their best during test (No)</i>			
Unknown			4.94*** (1.22)
Yes			5.43*** (1.04)
<i>Effort put into the test (Little or None)</i>			
Moderate			3.44** (1.55)
Lots			3.57** (1.63)
Best			3.96** (1.74)

Unknown			1.89 (1.92)
N	327	327	327
R2	0.08	0.08	0.25

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table R6: Test Score Regression Results, Combined Classes with Interactions, University

	Class	Characteristics	Faculty	Effort
(Intercept)	163.49*** (0.20)	160.94*** (0.64)	160.64*** (0.63)	150.17*** (1.14)
<i>Class (Entering)</i>				
Graduating	4.80*** (0.41)	6.06*** (1.36)	6.43*** (1.37)	0.79 (3.18)
<i>Sex (Female)</i>				
Male		1.88*** (0.42)	0.54 (0.44)	0.66 (0.41)
<i>High School GPA (< 3.50)</i>				
3.50-4.00		3.12*** (0.64)	2.20*** (0.64)	2.09*** (0.59)
I do not recall		1.30* (0.80)	0.88 (0.78)	1.07 (0.69)
I prefer not to respond		-0.41 (1.19)	-0.80 (1.14)	-1.34 (-1.14)
Unknown		2.23*** (0.75)	2.19*** (0.74)	2.34*** (0.69)
<i>Communicates better in English (No)</i>				
Yes		0.48 (0.44)	0.52 (0.43)	0.55 (0.41)
Equal		1.33* (0.77)	1.30* (0.74)	1.33** (0.69)
<i>Visible minority (No)</i>				
No response		-3.66 (2.36)	-5.12** (2.38)	-2.47 (2.69)
Yes		-2.54*** (0.47)	-2.52*** (0.47)	-2.29*** (0.45)
Unknown		-1.65** (0.70)	-1.34** (0.69)	-1.36** (0.65)
<i>Class (Entering), Sex (Female)</i>				
Graduating * Male		-1.13 (0.91)	0.4 (0.93)	1.33 (0.82)
<i>Class (Entering), High School GPA (< 3.50)</i>				
Graduating * 3.50-4.00		-0.08 (1.31)	0.86 (1.32)	1 (1.27)
Graduating * I do not recall		-0.85 (1.93)	-0.35 (1.92)	-0.34 (1.78)
Graduating * I prefer not to respond		-5.30 (4.13)	-4.98 (4.12)	-4.28 (4.37)
Graduating * Unknown		-0.92 (2.30)	-0.86 (2.30)	-0.66 (2.04)
<i>Class (Entering), Communicates better in English (No)</i>				

Graduating * Yes	-1.03	-1.06	-1.21
	-0.9	-0.9	-0.86
Graduating * Equal	-0.90	-0.95	-0.82
	-1.59	-1.57	-1.51
<i>Class (Entering), Visible minority (No)</i>			
Graduating * No response	0.38	1.98	0.05
	(3.96)	(3.96)	(3.72)
Graduating * Yes	-0.14	-0.18	-0.33
	(0.94)	(0.94)	(0.84)
Graduating * Unknown	1.8	1.49	2.01
	(2.65)	(2.66)	(2.20)
<i>Faculty (Arts & Science)</i>			
Engineering		3.36***	2.04***
		(0.46)	(0.82)
<i>Class (Entering), Faculty (Arts & Science)</i>			
Graduating * Engineering		-3.84***	-1.84*
		(0.92)	(1.11)
<i>Answered 75%+ of the Test (No)</i>			
Yes			7.73***
			(0.93)
<i>Tried their best during test (No)</i>			
Unknown			0.8
			(0.79)
Yes			0.59
			(0.53)
<i>Effort put into the test (Little or None)</i>			
Moderate			2.70***
			(0.54)
Lots			4.46***
			(0.66)
Best			5.96***
			(0.87)
Unknown			3.01***
			(0.97)
<i>Class (Entering), Answered 75%+ of the Test (No)</i>			
Graduating * Yes			0.19
			(2.43)
<i>Class (Entering), Tried their best during test (No)</i>			
Graduating * Unknown			4.14***
			(1.45)
Graduating * Yes			4.83***
			(1.17)
<i>Class (Entering), Effort put into the test (Little or None)</i>			
Graduating * Moderate			0.74
			(1.64)
Graduating * Lots			-0.89

				(1.76)
Graduating * Best				-2.00
				(1.95)
Graduating * Unknown				-1.12
				(2.15)
N	1402	1402	1402	1402
R2	0.09	0.17	0.2	0.32

*** p < 0.01; ** p < 0.05; * p < 0.1.



Higher Education
Quality Council
of Ontario

An agency of the Government of Ontario