Higher Education Quality Council of Ontario

An agency of the Government of Ontario

Opportunities and Challenges in Predictive Modelling for Student Retention Appendix

Patrick Lougheed, Andrew Drinkwater and Lynne Jamieson

Plaid Consulting

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:

Lougheed, P., Drinkwater, A. & Jamieson, L. (2018). *Opportunities and Challenges in Predictive Modelling for Student Retention Appendix.* Toronto: Higher Education Quality Council of Ontario.

Appendix A: Predictive Modelling

Predictive modelling is, fundamentally, the use of data to predict outcomes using statistical modelling techniques (Finlay, 2014). In other words, the idea is to predict future behaviour and/or outcomes based on what is known about the past (van Barneveld, Arnold & Campbell, 2012; Finlay, 2014). It uses techniques from several fields such as data mining, statistical modelling and machine learning to investigate relationships among data variables and develop prediction algorithms. Predictive modelling is neither a “magic bullet” nor a fully automated process — models are only as good as their underlying assumptions, and making logical, correct assumptions requires substantial time and energy on the part of a knowledgeable, experienced group of people with expertise in both the business (in this case, higher education, students and retention) and technical (data and statistics) domains. Predictive modelling is not synonymous with “big data,” a term that has gained cultural traction. Big data instead refers to data sets sufficiently large that they tax or are beyond the ingestion, storage, querying and management capabilities of traditional database systems (Manyika et al., 2011). Big data may be used in predictive modelling, but the terms refer to different systems and techniques.

“Learning analytics” is an often heard term in the higher education domain, and generally focuses on using measures of student engagement with learning material, most often gathered from a learning management system, or LMS, (e.g., Blackboard, Desire2Learn BrightSpace, Moodle, Minerva, etc.) to predict which students may have difficulty meeting curricular goals such as successful course completion. The broader concept of “academic analytics” (van Barneveld, Arnold & Campbell, 2012) can involve looking at longer-timeframe data and outcomes, such as degree completion, utilizing data gathered from multiple institutional systems. Predictive modelling can help the institution “achieve and maintain the optimum recruitment, retention and graduation rates of students” necessary to meet its strategic enrolment management goals. (Dolence, 1993, p. 8)

The process behind predictive modelling is usually similar between projects. The first step involves defining the outcome to be predicted and how that outcome is represented in past data. Available data is then gathered and analyzed for potential usefulness as inputs to a predictive model, with a particular eye to relationships between the possible inputs and the outcome to be measured. Data may require cleansing or shaping, accounting for how data is collected and the form it takes; as an example, an LMS may track each particular interaction a student has with a course, but the predictor data for the model may simply be the number of interactions, and the data point must be aggregated and shaped appropriately. Test models are created and assessed to see how well they predict the true outcomes, and ultimately one or a few models are chosen. Models are then constantly tested and refined — new data may be added as it becomes available, or changes to the context may mean certain data is no longer useful and must be removed or replaced. The factors that lead to students succeeding are varied, complex and often difficult to measure (Wiggers & Arnold, 2011; Richardson, Abraham & Bond, 2012), and the discovery or creation of alternative or proxy measures is sometimes required.

One of the primary assessment mechanisms for predictive models is “accuracy,” specifically how many of the outcomes — in this context, examples are whether a student was actually retained, or whether a student successfully completed a course — where correctly predicted. Works by Delen (2010) and Zhang, Oussena, Clark and Kim (2010) suggested that 80% accuracy is achievable with sufficient data and appropriate techniques. While the survey conducted as part of this report did not explicitly ask respondents how they defined accuracy, a common approach is to split students into two groups: for example, when attempting to predict whether a student would be retained, these groups would be “retained” and “not retained.” Where the predictive model accurately predicts the outcome of retained or not retained, this is considered accurate. False positives, where the model predicts retained but the actual result is not retained; and false negatives, where the model predicts not retained but the actual result is retained, are considered to be errors on the part of the model.

Accuracy in prediction is important, as the goal of prediction is to do better than the assumption that everyone will be retained. For example, if an institution with a 75% retention rate assumed everyone would be retained, they would be accurate 75% of the time, including 100% of the time for students who were retained and 0% of the time for those that were not. Embedded within the accuracy of the system is the notion of both false-positives and false-negatives. A false-positive predicts a student being retained even though they are not retained, which may impact the institution’s decision to provide supports to that student. Conversely, a false-negative predicts non-retention when the student is retained, which may divert student support resources toward a student who will otherwise succeed on their own. If the predictive model can’t improve upon the “everyone is retained” assumption or the model has a substantial number of false positives or false negatives, then the institution is better off continuing with current processes rather than working with students in different ways depending on whether they are predicted to be retained or not.

Delen’s (2010) research noted a difference in accuracy between classes, with the retained group accurately predicting students who would be retained with greater than 90% accuracy using four different models. However, the prediction accuracy dropped substantially to less than 50% for the not retained group, which was deemed not acceptable. The author attributed this to the skewness of the original data set (which was approximately 80% retained and 20% not retained); similar retention rates are common at many institutions. These difficulties in predicting the not retained group correctly can make it difficult to ensure individual-level proactive outreach reaches the right individuals. To get past this accuracy challenge, the author recommends taking the entire minority group (not retained) and randomly sampling the majority group (retained) so that it will be the same size — this resulted in a sizable lift in accuracy of between 74% and 80%. Ensemble models improved this result incrementally, increasing accuracy for the not retained group to between 80% and 82% (Delen, 2010).

Appendix B: Glossary of Terms

Classification: Classifications group and organize information meaningfully and systematically into a standard format that is useful for determining the similarity of ideas, events, objects or persons. (Hoffmann & Chamie, 2002).

Customer relationship management (CRM) System: A system that tracks interactions with customers (students), often from prior to the time of application through to graduation. CRM systems in higher education often track electronic, telephone and in-person communications with prospective and current students as well as alumni, supporting the recruitment, enrolment and retention goals of the institution.

Data mart: A subset of a data warehouse, often limited to a single subject area.

Data warehouse: A historical repository of data used to support the decision-making process throughout the organization. A data warehouse spans multiple subject domains and provides a consistent view of data objects used by various business processes throughout the enterprise (Horvli, 2004).

Enrolment management: "An organizational concept and systematic set of activities designed to enable educational institutions to exert more influence on their student enrolments" (Hossler & Bean, 1990, p. 5).

Learning management system (LMS): "A software application that automates the administration, tracking, and reporting of training events" (Ellis, 2009, p.1). An LMS will typically contain information on student interactions with course materials that reside on the LMS, such as which materials were read, when and for how long; which assessments were attempted and when; and the assessment results. Examples of common learning management systems used by Canadian institutions include BlackBoard's Learn, D2L's BrightSpace, Instructure's Canvas and Moodle.

Machine learning: "The branch of computer science that utilizes past experience to learn from and use its knowledge to make future decisions" (Dangeti, 2017, p.8).

Predictive modelling: Predictive modelling is the application of statistical and informational modelling techniques such as classification, regression and machine learning to make predictions based on previously recorded observations (Finlay, 2014).

Regression: Identifies the relationship between independent and dependent variables (Reinard, 2006).

Retention: Within the context of this project, student retention refers to any measure of student enrolment or performance past the point of first enrolment, such as: year-to-year persistence; graduation; performance as measured by average grades; performance as measured by rate of good academic standing; or performance in individual courses.

Strategic enrolment management (SEM): "A comprehensive process designed to help an institution achieve and maintain the optimum recruitment, retention and graduation rates of students, where optimum is defined in the academic context of the institution" (Dolence, 1993, p. 8).

Student information system (SIS): A system that manages student data such as course enrolments, financial aid, admission and graduation. An SIS of a postsecondary institution generally contains similar information to what would appear on the official academic records: identifying information such as name, biographic and demographic information, contact information and identifying numbers; previous educational history, including schools and relevant courses; programs enrolled in; courses enrolled in at the institution with final grades and notations; academic standings; academic honours and awards; and graduation records. Additionally, many student information systems contain admissions information such as evaluations; financial aid information such as applications for and awarding scholarships, awards and bursaries, and confirmations of eligibility for government aid; and financial accounts such as tuition and other fees charged and paid. Examples of common student information systems used in Canadian institutions include Ellucian's Banner and Colleague products, and Oracle's PeopleSoft Campus Solutions.

Transfer: "Students who begin studies at one institution and later transfer to another" (State University, 2018, Defining Student Retention section, para. 1).

Underrepresented groups: Groups of students who are not representative of their proportion of the wider population. Some underrepresented groups identified by prior research include: low-income students, first-generation PSE students, rural students, first- and second-generation children of immigrants, Aboriginal Canadians, French speakers outside Quebec, students from single-parent families and students with a disability (Finnie, Childs & Wismer, 2011).

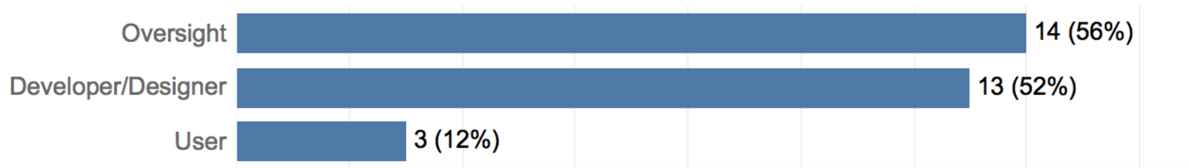
Appendix C: Additional Results — Phases One and Two

This appendix contains the complete results of the phase one survey questions and phase two interviews/questionnaires that are not discussed in detail in the results section of the paper.

Phase One: Survey

Question 6: This question asked respondents about their role with regards to predictive modelling. Respondents were asked to select all that applied. Most of the survey responses came from individuals with either oversight responsibility (14 responses, 56%) or a role involved in developing the predictive model (13 responses, 52%), with three responses (12%) from users of predictive models. Some respondents held more than one role. See Figure C1 for this breakdown.

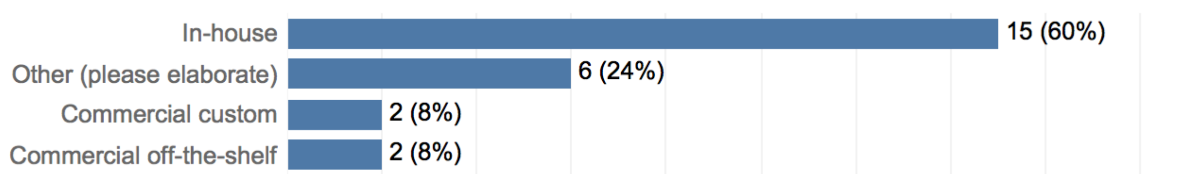
Figure 1: This appendix contains the complete results of the phase one survey questions and phase two interviews/questionnaires that are not discussed in detail in the results section of the paper.



Question 8: This question asked what type of system is being used for predictive modelling, one developed in-house, commercially available, or other (see Figure C2). A majority of respondents (15, 60%) indicated they were using a system developed in-house, while two (8%) were using a system available commercially off-the-shelf and two others (8%) were using a commercial system customized for their needs. In the "other" group (six respondents, 24%) two respondents indicated they used a mixture of commercial and in-house systems.

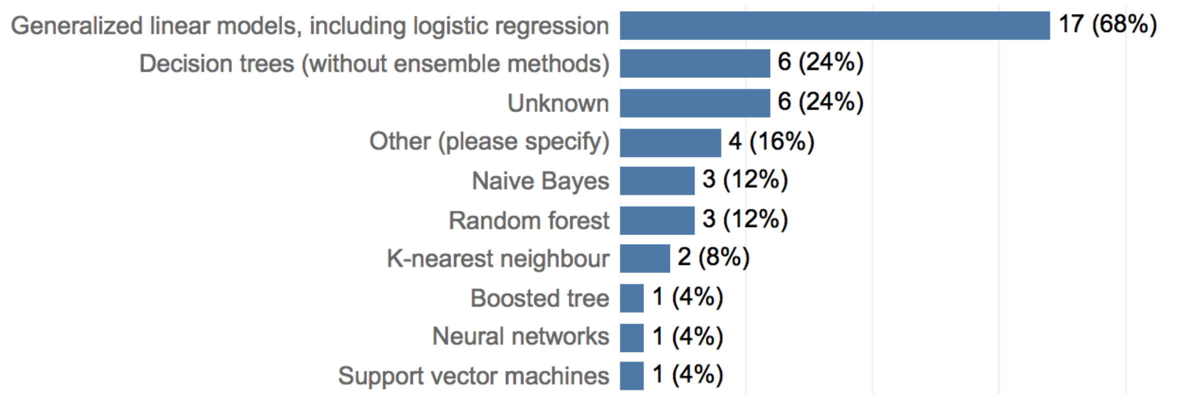
The commercially available products mentioned by respondents included EAB, Pharos360 and SkyFactor's MapWorks. Others mentioned include RapidInsight, SAS and IBM's SPSS1, which while providing predictive modelling off-the-shelf, are not customized to student retention models and some development of the models need to be undertaken. SPSS was also mentioned in the "other" category, as were systems such as Python and R, which similarly provide modelling functionality but not retention models; in-house models were largely built with the same kinds of technology.

Figure 2: Responses to Question 8, "Are you using a commercial off-the-shelf system for your predictive modelling for student retention, a system developed in-house or something else?" (Select One)



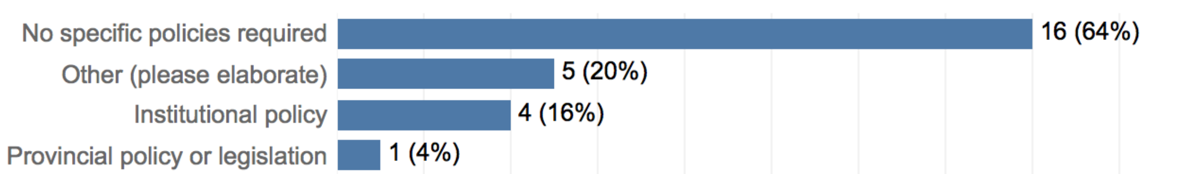
Question 10: This question asked about the predictive modelling techniques being used. As shown in Figure C3, the most commonly used technique was generalized linear models with 17 responses (68%). Some respondents (six, 24%) were not aware of the specifics of their system, and an additional respondent did not select any of the potential answers to this question. Decision trees (six, 24%) were the only other response that garnered more than three responses. In the "other" category, techniques used included Markov chains and elaborated linear models, while the other two responses did not provide specific techniques.

Figure 3: Responses to Question 10, “What predictive modelling techniques are you using for student retention?” (Select All That Apply)



Question 15: This question asked about specific policies or legislation governing the use of predictive modelling for student retention. Most respondents (16, 64%) indicated that there were no specific policies that governed their use of predictive modelling for student retention (Figure C4), while four (16%) indicated there were institutional policies. One respondent (4%) selected provincial policy or legislation, while the “other” category (five, 20%) included responses indicating their institutional or other policies that applied to student privacy more generally; another response to “other” pointed to the institutional strategic enrolment management plan.

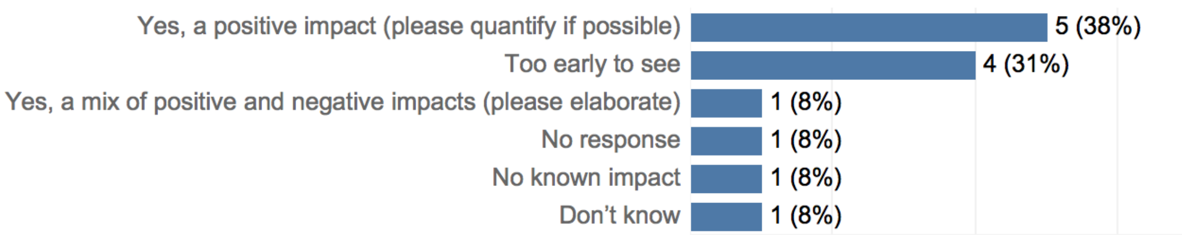
Figure 4: Responses to Question 15, “Are there specific policies or legislation governing your use of predictive modelling for student retention?” (Select All That Apply)



Questions 22 to 25: These questions asked about the impacts of predictive modelling on student success measures such as retention/persistence rates, academic standing, graduation/completion rates and student performance.

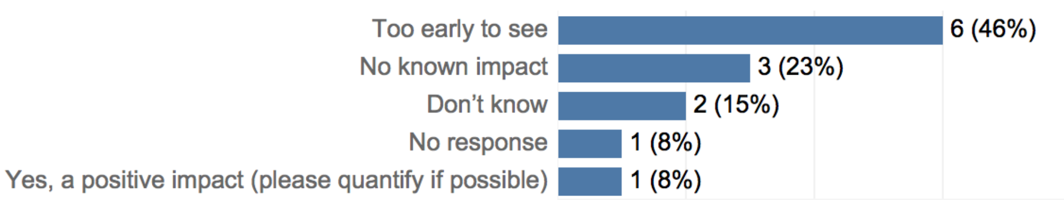
For retention and persistence rates (Figure C5), five respondents (38%) indicated that predictive modelling, coupled with interventions, had a positive impact, while a further four respondents (31%) indicated it was too early to see, and one respondent (8%) indicated a mix of positive and negative impacts. One respondent in the positive group indicated increases in retention rates of approximately 5%, while the remainder could not provide specific numbers. The respondent indicating a mix of positive and negative noted that no systematic analysis had been conducted, but that anecdotally the models had flagged some students who would have been missed through other methods as well as some students who did not require interventions.

Figure 5: Responses to Question 22, "Has your use of predictive modelling had an impact on retention/persistence rates?" (Select One)



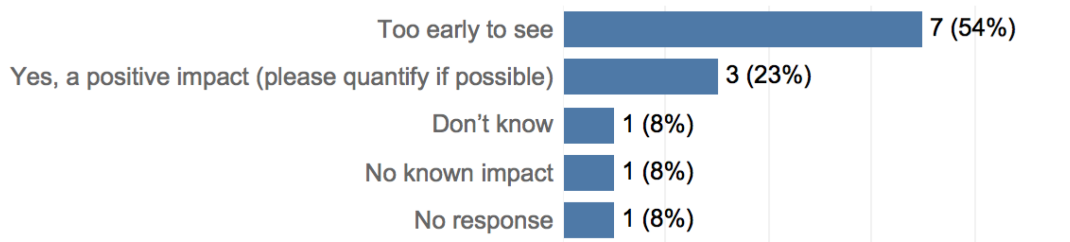
For academic standing rates (Figure C6), six respondents indicated it was too early to see (46%), three said that there was no known impact (23%) and two that they did not know (15%). Among the remaining responses indicating a positive impact (one, 8%) the respondent framed it as "continued improvement."

Figure 6: Responses to Question 23, "Has your use of predictive modelling had an impact on the rate of students being in unsatisfactory academic standing (such as on probation, required to withdraw, dismissed, etc.)?" (Select One)



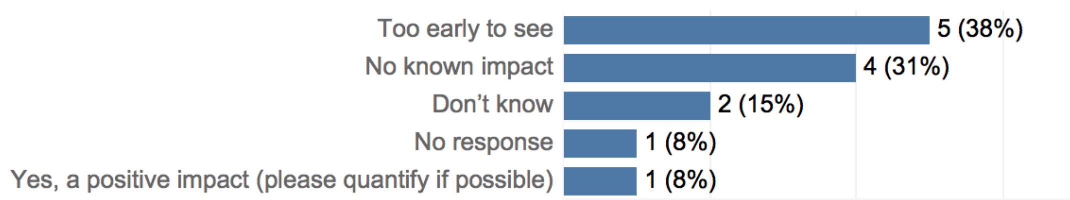
For graduation/completion rates (Figure C7), more respondents indicated that it was too early to see the effect of predictive modelling (seven, 54%), while smaller numbers reported no known impact (one, 8%) or that they did not know (one, 8%). Among three respondents (23%) who indicated a positive impact, one respondent indicated a 2% improvement, one a 5–10% improvement each year over several years, and one indicated that there were too many variables to identify the specific impact of predictive modelling and interventions.

Figure 7: Responses to Question 24, "Has your use of predictive modelling had an impact on graduation/completion rates?" (Select One)



Finally, when asked about the issue of student performance through the lens of grade point average or similar measures, fewer respondents (as compared to Questions 23 and 24) indicated that it was too early to see (five, 38%) and more responded that there was no known impact (four, 31%) or that they did not know (two, 15%). See Figure C-8.

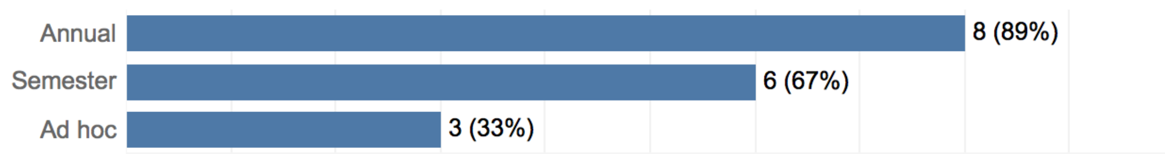
Figure 8: Responses to Question 25, "Has your use of predictive modelling had an impact on student performance, such as grade point averages?" (Select One)



Phase Two: Interview and Questionnaire

Predictive modelling for the purposes of enrolment and budget planning was more generally done on an annual basis as reported by eight of nine participants (89%), while modelling predicting retention outcomes was done on either every semester (six, 67%) or on an ad hoc basis as information became available (three, 33%). Information availability might be early-alert processes triggered by faculty or staff, regular data pulls from information systems, as data is collected and introduced to the model, or as new models are created for different predictive purposes. See Figure C9 for a breakdown of responses.

Figure 9: Participants' Scheduling of Predictive Modelling



Participants reported having models that were both aggregate — that is, predicted retention for a group of students without predicting retention for any particular student — and individual, as shown in Figure C10. There was significant overlap between these groups, as 3 participants (33%) reported using both individual and aggregate models; three participants (33%) reported having only individual models while three participants (33%) reported having only aggregate models. Models used for enrolment planning were all in aggregate, though one participant noted that they brought in individual student information to connect with the aggregate information for course planning purposes. Models used for retention prediction could be either aggregate or individual.

Figure 10: Participants' Aggregation of Predictive Modelling

Appendix D: Survey Instrument

Note: Questions marked with an * were mandatory responses if a respondent was shown that question; Question 34, for example, was skipped over if the respondent indicated in Question 4 that they did not use predictive modelling at their institution.

Preamble

The Higher Education Quality Council of Ontario (HEQCO) is looking to better understand whether, where and how predictive modelling is being used effectively by postsecondary institutions to improve student retention. For the purposes of this research, we're using the term "predictive modelling" very broadly to mean: the application of statistical and informational modelling techniques such as classification, regression, and machine learning to make predictions based on previously recorded observations. Other terms used include "predictive analytics" and "learning analytics." We are also using a wide lens on "student retention" to capture uses of modelling for any predictive purposes past the point of first enrolment, such as: year-to-year persistence; graduation; performance as measured by average grades; performance as measured by rate of good academic standing; or performance in individual courses.

This survey focuses on the use of predictive modelling for student retention across higher education, how interventions are informed and measured, what types of policies were developed or adhered to in relation to predictive modelling and challenges faced by institutions.

The survey will take approximately 25 minutes, depending on the length of your answers, and your progress can be saved so that you can return to the survey and finish later. To save your progress, press the "Save Page and Continue Later" button at the bottom of any page; you will be asked for your email address and the system will send you a link you can use to return to the survey.

Your ideas and those of other stakeholders from postsecondary institutions will help in the creation of predictive modelling tools for student retention and interventions informed by predictive modelling.

Participation is voluntary. Neither you nor your institution will be identified in the report. All of the information that you share will remain anonymous for the final report. This means that only researchers at Plaid Consulting and HEQCO will have access to identifying information in connection with survey findings, but that identifying information will be aggregated or anonymized prior to publication. More information on the privacy policy can be found at http://plaid.is/privacy.html. Any feedback that you provide will remain confidential in accordance with Canada's Freedom of Information and Protection of Privacy Act.

This survey is being conducted by Plaid Consulting on behalf of HEQCO. If you have any questions about the survey, please contact Pat Lougheed, Partner and Co-Founder, Plaid Consulting, at info@plaid.is or +1-604-306-0199. If you have any questions about the project, please contact Kaitlyn Blair, Researcher at HEQCO, at kblair@heqco.ca or +1-416-212-3881.

Section 1: Identification

* 1. Please identify your institution: __________________
2. Please identify the location of your institution: __________________
3. What is your title at your institution? __________________

Section 2: Modelling System and Inputs

* 4. Is your institution currently using predictive modelling for student retention purposes?

We are using "student retention" to capture uses of modelling for any predictive purposes past the point of first enrolment, such as: year-to-year persistence; graduation; performance as measured by average grades; performance as measured by rate of good academic standing; or performance in individual courses.

a. ( ) Yes
b. ( ) Not currently using predictive modelling, but are planning to implement soon, or are investigating or seriously considering implementing
c. ( ) Not currently using predictive modelling, but have looked into it in the past

d. ( ) Not currently using predictive modelling, and have not considered it
e. ( ) Not currently using predictive modelling, but have used it in the past

If response is "Not currently using predictive modelling, but have used it in the past" then display comment box: "Are you able to elaborate on why you stopped using predictive modelling?"

Otherwise, if response is not yes, display comment box: "Are you able to elaborate on why you're not currently using predictive modelling, whether you think this may change in the future?"

"Yes" continues the survey; others are directed to thank you page.

5. You indicated that your predictive model is used for student retention purposes. In what ways does your institution use predictive modelling? Please select all that apply:

a. [ ] Identifying students at risk of leaving for academic performance reasons
b. [ ] Identifying students at risk of leaving for mental health reasons
c. [ ] Identifying students at risk of leaving due to a disability
d. [ ] Identifying students at risk of leaving for financial reasons
e. [ ] Targeting interventions toward students at risk of leaving
f. [ ] Promoting the use of academic and/or advising resources
g. [ ] Determining which interventions improve student retention
h. [ ] Determining effective admissions criteria
i. [ ] Designing more effective curriculums
j. [ ] Improving enrolment planning (i.e., predicting which students will continue at the institution)
k. [ ] Other (please elaborate: __________)

6. What is your role with regards to predictive modelling for student retention at your institution?
____

7. When did you begin using predictive modelling for student retention purposes? Select one:

a. ( ) 2017
b. ( ) 2016
c. ( ) 2015
d. ( ) 2014
e. ( ) 2013
f. ( ) 2012
g. ( ) 2011
h. ( ) 2010
i. ( ) 2009
j. ( ) 2008
k. ( ) 2007
l. ( ) 2006

m. ( ) 2005
n. ( ) 2004
o. ( ) 2003
p. ( ) 2002
q. ( ) 2001
r. ( ) 2000
s. ( ) Prior to 2000
t. ( ) Unsure

8. Are you using a commercial off-the-shelf system for your predictive modelling for student retention, a system developed in-house or something else? Select one:

a. ( ) Commercial off-the-shelf
b. ( ) Commercial custom
c. ( ) In-house
d. ( ) Other (please elaborate: _________________)

9. When you originally implemented your predictive modelling for student retention, who was involved in the implementation? Select all that apply:

a. [ ] In-house staff
b. [ ] In-house faculty
c. [ ] System vendor
d. [ ] External consultants or other organization (please identify: _________________)

10. What predictive modelling techniques are you using for student retention? Select all that apply:

a. [ ] Naive Bayes
b. [ ] K-nearest neighbour
c. [ ] Decision trees (without ensemble methods)
d. [ ] Random forest
e. [ ] Boosted tree
f. [ ] Support vector machines
g. [ ] Neural networks
h. [ ] Generalized linear models, including logistic regression
i. [ ] Unknown
j. [ ] Other (please elaborate: _________________)

11. What systems do your predictive models for student retention gather information from? Select all that apply and please identify the system vendor below:

a. [ ] Student Information System (SIS) (Which SIS vendor?: _________________)
b. [ ] Learning Management System (LMS) (Which LMS vendor?: _________________)
c. [ ] Advising System (if separate from SIS) (Which advising system vendor?: _________________)
d. [ ] Financial Aid System (if separate from SIS) (Which financial aid system vendor?: _________________)
e. [ ] Student Engagement Tracking System (if separate from SIS) (Which student engagement tracking system vendor?: _________________)
f. [ ] Other (Which other systems?: _________________)

12. What types of information do your predictive models use for student retention? Select all that apply:

a. [ ] Demographic (examples: gender, age, race, ethnicity, citizenship)
b. [ ] Language abilities (examples: proficiency in the institution's primary language of instruction, proficiency in other languages)
c. [ ] Location (examples: home distance from campus, living distance from campus, in residence)
d. [ ] Previous educational history at your institution
e. [ ] Previous educational history in secondary/high school
f. [ ] Previous educational history at other postsecondary/tertiary education institutions (where applicable)
g. [ ] Learning management system interactions
h. [ ] On-campus activities
i. [ ] Self-assessment questionnaires (Are you able to tell us a little more about the kinds of self-assessment questionnaires you use?: _________________)
j. [ ] Standardized test scores (Which standardized tests do you use?: _________________)
k. [ ] Other (please elaborate: _________________)

13. For which student populations does your institution use predictive models for student retention? Select all that apply:

a. [ ] All students
b. [ ] All undergraduate students
c. [ ] All graduate students
d. [ ] Professional degree (law, medicine, etc.) students
e. [ ] First-year/freshmen undergraduate students
f. [ ] Direct entry from high school students
g. [ ] Transfer students from other postsecondary institutions
h. [ ] Mature students
i. [ ] Indigenous students
j. [ ] First-generation students
k. [ ] Low-income students

l. [ ] Distance-education students
m. [ ] Students with disabilities
n. [ ] Students in particular faculties/schools/colleges (You selected 'Students in particular faculties/schools/colleges' above. Please identify them here: _________________)
o. [ ] Other (please elaborate: _________________)

14. What was the impetus for your use of predictive modelling for student retention? Select all that apply:

a. [ ] Federal or provincial requirements or priorities
b. [ ] Institutional requirements or priorities
c. [ ] Student success reasons
d. [ ] Budgetary reasons
e. [ ] Other (please elaborate: _________________)

15. Are there specific policies or legislation governing your use of predictive modelling for student retention? Select all that apply:

1. [ ] Federal policy or legislation
2. [ ] Provincial policy or legislation
3. [ ] Institutional policy
4. [ ] No specific policies required
5. [ ] Other (please elaborate and provide links where possible: _________________)

Section 3: Modelling Outcomes

16. Who has access to the predictions from the predictive modelling system? Select all that apply:

a. [ ] Academic advisors
b. [ ] Non-academic advisors (for example, counsellors)
c. [ ] Faculty members
d. [ ] Non-faculty instructors
e. [ ] Administrative managers
f. [ ] Administrative staff
g. [ ] Other (please identify: _________________)

17. How are the predictions available to those with access? Select all that apply:

a. [ ] Within the predictive modelling system only
b. [ ] Within a data mart/data warehouse with reports
c. [ ] Within the student information system
d. [ ] Within the learning management system
e. [ ] Within the advising system
f. [ ] Other (please identify: _________________)

18. Do you find that your current predictive modelling system and methods work well for your needs? If not, why not? Select one:

a. ( ) Yes
b. ( ) Don't know
c. ( ) No (please elaborate: ________________)

19. Please indicate what percentage of students are predicted accurately in your modelling:

	0–49%	50–59%	60–69%	70–79%	80–89%	90–100%	Unsure
All students							
All undergraduate students							
All graduate students							
Professional degree (law, medicine, etc.) students							
First-year/freshmen undergraduate students							
Direct entry from high school students							
Transfer students from other postsecondary institutions							
Mature students							
Indigenous students							
First-generation students							
Low-income students							
Distance-education students							
Students with disabilities							
Students in particular faculties/schools/colleges							
Other (as you identified in Q13)							

Respondents were asked to select one category for each group they had identified as being included in their predictive modelling in Question 13. Groups not selected in Question 13 were not shown.

Section 4: Interventions

20. Do you currently use predictive modelling to inform specific student retention interventions? Select one:

a. ( ) Yes
b. ( ) No (please elaborate: _________________)

If no, jump to Q26.

21. Please indicate the different types of interventions your institution uses based on your predictive models. Select all that apply:

a. [ ] Promotion of available support services
b. [ ] Access to self-assessment tools (such as a learning skills assessment)
c. [ ] Optional individual advising
d. [ ] Mandatory individual advising
e. [ ] Optional group advising
f. [ ] Mandatory group advising
g. [ ] Optional mentoring
h. [ ] Mandatory mentoring
i. [ ] Optional educational scaffolding (such as an “introduction to postsecondary studies” course)
j. [ ] Mandatory education scaffolding (such as an “introduction to postsecondary studies” course)
k. [ ] Other (please identify: _________________)

22. Has your use of predictive modelling had an impact on retention/persistence rates? Select one:

a. ( ) Yes, a positive impact (Please quantify the positive impact: _________________)
b. ( ) Yes, a negative impact (Please quantify the negative impact: _________________)
c. ( ) Yes, a mix of positive and negative impacts (Please elaborate on the mix of positive and negative impacts: _________________)
d. ( ) No known impact
e. ( ) Too early to see
f. ( ) Don’t know

23. Has your use of predictive modelling had an impact on the rate of students being in unsatisfactory academic standing (such as on probation, required to withdraw, dismissed, etc.)? Select one:

a. ( ) Yes, a positive impact (Please quantify the positive impact: _________________)
b. ( ) Yes, a negative impact (Please quantify the negative impact: _________________)

c. ( ) Yes, a mix of positive and negative impacts (Please elaborate on the mix of positive and negative impacts: _________________)
d. ( ) No known impact
e. ( ) Too early to see
f. ( ) Don't know

24. Has your use of predictive modelling had an impact on graduation/completion rates? Select one:

a. ( ) Yes, a positive impact (Please quantify the positive impact: _________________)
b. ( ) Yes, a negative impact (Please quantify the negative impact: _________________)
c. ( ) Yes, a mix of positive and negative impacts (Please elaborate on the mix of positive and negative impacts: _________________)
d. ( ) No known impact
e. ( ) Too early to see
f. ( ) Don't know

25. Has your use of predictive modelling had an impact on student performance, such as grade point averages? Select one:

g. ( ) Yes, a positive impact (Please quantify the positive impact: _________________)
h. ( ) Yes, a negative impact (Please quantify the negative impact: _________________)
i. ( ) Yes, a mix of positive and negative impacts (Please elaborate on the mix of positive and negative impacts: _________________)
j. ( ) No known impact
k. ( ) Too early to see
l. ( ) Don't know

Section 5: Success

26. What are the biggest challenges you've faced related to modelling student retention? _________________

27. What are the biggest successes you've had with modelling student retention? _________________

28. What advice would you offer to an institution looking to implement predictive modelling for student retention? _________________

29. Are there types of information you would like to use in your predictive modelling but have not yet been able to integrate? Select one:

a. ( ) Yes (Please elaborate on data types and reasons why not: _________________)
b. ( ) No

30. Has your predictive modelling system been reviewed since it was originally implemented to see if refinements can be made? Select one:

a. ( ) Yes, this is done annually
b. ( ) Yes, this is done regularly (every few years)
c. ( ) No

If no, jump to Question 32.

31. Is this model review an in-house process, or do you work with an external organization? Select all that apply:

a. [ ] In-house staff
b. [ ] In-house faculty
c. [ ] System vendor
d. [ ] External consultants or other organization

32. Has the use of predictive modeling at your institution lead to any major changes in the following areas? Please select all that apply:

a. [ ] Curriculum design
b. [ ] General education options
c. [ ] Educational scaffolding courses (such as an "introduction to postsecondary studies" course)
d. [ ] Academic advising
e. [ ] Promoting of academic resources
f. [ ] Criteria for admission into your institution
g. [ ] Criteria for admissions into specific programs offered at your institution
h. [ ] Student retention policies
i. [ ] Other (please elaborate: ___________)
j. [ ] Predictive modelling has not led to major changes at our institution

33. Have you investigated alternative methods of modelling student retention, such as using a different system or different predictive methods? Select one:

a. ( ) Yes, in process of investigating
b. ( ) Yes, and changed to a new method of modelling (What methodology did you change from and to, and why?: ___________)
c. ( ) Yes, but opted to stay with the current system/method
d. ( ) No

Section 6: Followup

* 34. Would you be willing to answer followup questions by email and/or participate in the phone interview portion of the project?

Followup questions will generally relate explicitly to your responses today — for example, where we need clarification to make sure we fully understand your response. The interview portion will delve deeper into your experiences with predictive modelling.

a. ( ) Followup questions by email only
b. ( ) Interview only
c. ( ) Followup questions and/or interview
d. ( ) Please do not follow up with me

If "Please do not follow up with me," jump to thank you page.

35. Your Contact Information
- First Name:
- Last Name:
- Phone:
- Email Address:

We recognize that at some institutions, multiple people work with predictive modelling. Is there someone else at your institution who you would recommend we reach out to for additional information?

- First Name:
- Last Name:
- Phone:
- Email Address:

Appendix E: Interview and Email Questionnaire Instrument

Context of Predictive Modelling

1. Please describe how predictive modelling is used at your institution. If it is used within a specific department or smaller unit, please describe the use in the specific context. (i.e., institution-wide, department-wide, etc.).
2. How often is the predictive modelling done (i.e., daily, weekly, monthly, by semester, etc.)?
3. Is your model done in aggregate (i.e., of 100 people in group xyz, 90 will retain) or individually (i.e., student John Doe has an 80% likelihood of retaining)?

Motivations Behind Your Use of Predictive Modelling

4. Please describe why your institution and/or your department decided to use predictive modelling.
5. Where did the request to use predictive modelling originate (i.e., president's office, etc.)?

Implementation of Predictive Modelling and Use of Data

6. In the use of predictive modelling that you described earlier, how, if at all, has use evolved the model's initial design?
7. What data sources are used within the predictive model? For example, do you rely only on institutional data or do you use external data sources (i.e., application data from OUAC/OCASS (in Ontario), school-board-specific data, resources from Statistics Canada, etc.)?
8. How do you ensure that the predictive model is using the data in the way the data was intended to be used?

Challenges with Predictive Modelling

9. Please describe any challenges that you have encountered (i.e., staff resources, funding, unable to interpret the results, etc.).
   - If you identified challenges above, how were you able to overcome these issues?

10. Please describe anything that you did not take into account that you wished you had considered when you started (i.e., timelines, comparability between data sets, resources, etc.)?

Benefits of Predictive Modelling and Next Steps

11. Did any changes result from your use of the predictive model (i.e., specific interventions, institutional policy changes, etc.)? If so, please describe these specific changes.
12. If you answered YES to Question 11, how has the data from your predictive model been used to inform intervention development?
13. What has been most helpful in transitioning from a theoretical tool to an applied context?
14. In the next five years, how, if at all, will your institution continue to use predictive modelling?
15. Is there anything else you would like to comment on?

Higher Education Quality Council of Ontario

An agency of the Government of Ontario