Opportunities and Challenges in Predictive Modelling for Student Retention

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.* Toronto: Higher Education Quality Council of Ontario.

Higher Education Quality Council of Ontario

Acknowledgements

This research has been supported by dozens of stakeholders across Canada and around the world. We would like to express our sincerest thanks to the many who have been involved in the process.

First, to the respondents from several institutions who took the time to share their experiences through an online survey, and to those who agreed to follow up with an interview or an email. Your insight has been instrumental in helping us better understand the landscape of predictive modelling for student retention in higher education.

Second, to the industry groups that helped us ensure that the message about this research was shared globally, which helped us to attract respondents with deep knowledge of the technical underpinnings, usage and future plans for their models. This included the Association of Registrars of the Universities and Colleges of Canada (ARUCC); the Canadian Association of University and College Student Services (CACUSS); the Canadian Institutional Research and Planning Association (CIRPA); the Australia and New Zealand Student Services Association (ANZSSA); AMOSSHE, the Student Services Organization in the United Kingdom; the Association for Institutional Research (AIR); and the American Association of College Registrars and Admissions Officers (AACRAO). We also appreciate the efforts of many on social media platforms who engaged with our posts in order to help us reach a wider audience, and to all those who referred us other potential respondents.

Third, to the Higher Education Quality Council of Ontario (HEQCO) for funding this research and supporting us on the journey. We have greatly appreciated the support of Kate Hagerman and Lauren Hudak, who have worked with us in-depth providing feedback and suggestions for improvement, and Cassandra Cao and Helen Tewolde who were involved in the initial stages of the project.

Executive Summary

Project Overview

Canada's population is becoming increasingly well-educated, as evidenced by rising postsecondary enrolment and attainment of postsecondary credentials (Statistics Canada, 2017). However, access, retention and ultimately success for students in underrepresented groups continue to pose a challenge. Identifying students at risk of not succeeding and scaling interventions to provide useful supports to those students is necessary. One approach to addressing this problem is the use of predictive statistical models (Finnie, Fricker, Bozkurt, Poirier & Pavlic, 2017). Predictive modelling, generally, 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).

Within higher education, predictive modelling can be used for enrolment management, improving student success indicators (e.g., program graduation, retention, GPA, etc.), fundraising and to inform many other outcomes. Predictive modelling typically makes use of data from learning management systems, student information systems and student surveys. Many predictive modelling projects have focused on single institutions, but there are more recent examples that look at models utilizing data from multiple institutions. In the Canadian context, Mohawk College implemented a predictive modelling system in 2012 (Finnie et al., 2017a), as have several other institutions surveyed for this study.

As an exploratory study, the primary purpose of this project was to provide an overview of the landscape of predictive modelling in Canada, not a synopsis of technical strategies on how to use predictive models.[1] This report focuses on the use of predictive modelling, illuminating whether, where and how predictive modelling is being used effectively to improve student success and retention.[2]

Through targeted outreach to postsecondary institutions, the departments and individuals most likely to use predictive modelling were identified for participation in the online survey. Some individuals also opted to participate in followup, in-depth, semi-structured interviews and/or questionnaires (as selected by the respondent) where they could provide more detail on their use, context and challenges as well as any actions flowing from their use of predictive models.

A two-pronged approach, including requests sent via the mailing lists of higher education industry groups and targeted emails to key stakeholders at Canadian institutions, was used to select institutions to receive the survey. Overall, 170 responses to the online survey were collected. Of these, 100 were excluded because no information was provided, or because test responses or duplicates of responses already were received;

1 Some technical insights on the use of predictive modelling are offered in Appendix A and Appendix C.
2 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.

international responses were also excluded because the small number of responses made meaningful analysis impossible. This left us with 70 responses for further analysis — 66 of which were complete, and four of which were partially complete. Of the 66 complete responses, 38 were from universities, and 28 were from colleges and polytechnic institutions.

Following the survey, respondents were asked if they would be willing to participate in either a followup interview via telephone or a questionnaire over email. Seven interviews and two email questionnaires were completed.

The research questions for this project were:

a. Which practices and/or principles are used in predictive modelling in the postsecondary context?
b. What are innovative uses of predictive models that influence student retention?
c. What opportunities and challenges are associated with interventions informed by predictive models?
d. How can postsecondary institutions leverage the strengths of predictive modelling to improve student retention? Have the results of predictive modelling changed institutional practices or policies and is there data showing whether these changes have been effective in increasing student access, retention or success?

Key Findings

Key findings include:

- 36% of respondents indicated that their institution was using predictive modelling for student retention; 39% indicated that their institution was investigating, seriously considering, or planning on using it; 10% were not using predictive modelling and had not considered it; and 16% had looked into predictive modelling in the past, but were not currently using it and had no plans to begin using it.
- Of those who were not using predictive modelling, the respondents noted that resourcing was an issue, either in terms of people, time or tools, while other respondents identified issues with data quality or understanding.
- 52% of respondents used predictive modelling to inform specific student-retention interventions. Of these, the most common interventions employed promotion of support services and optional individual advising.
- Some respondents used predictive modelling in connection with their strategic enrolment management plan.
- The following were among the most innovative uses of predictive modelling:
    - One institution used a predictive model to help inform which applicants would be offered a place in residence.

    - Two institutions said they made the prediction results available to students, coupled with information that allowed the students to determine their own path forward by choosing among supports available to them.
    - One institution reported involving their strategic enrolment management committee in the oversight process for their predictive model. This review allows the committee to view variables chosen for the model, their relative weights and accuracy, and to discuss data points that could be considered for inclusion in future.

- Respondents agreed on two major themes:
    - Predictive models must be resourced appropriately, both in terms of human capital and technical infrastructure.
    - Predictive models must be appropriately secured — limiting access to those who need to know for their roles in running or assessing the models, or coordinating interventions. These predictions should be treated as any other sensitive personal information would be.
- Challenges experienced with modelling include data availability, transforming and mapping data as well as institutional acceptance and resourcing issues.
- Successes experienced with modelling include a shift in culture and conversation, as well as new or improved interventions and supports.
- The most common advice for those considering predictive modelling was to focus on communications and buy-in, as well as integration and communication between areas of the institution (e.g., faculties, academic programs, administrative departments).
- Due in part to budgetary pressures and an increasing need to support student success, the majority of predictive modelling systems have been implemented in the last four years, and are often built in-house.

Conclusions and Future Research Directions

Informing intervention activities on the basis of predictive modelling is a challenge — 52% of respondents that have a predictive model reported using it alongside existing interventions. The trajectory of predictive modelling, particularly in Canadian higher education, is still in its early days, with several respondents suggesting that interventions were on their road map for the next couple of years. Others suggested it was challenging to get those who are responsible for interventions to see value in the modelling. One of the biggest concerns brought up by respondents was around resourcing, in terms of time, people and infrastructure.

While many respondents felt that it was too early to see the impact of predictive modelling, those who had seen an impact reported it was a positive one. A specific positive impact that many respondents noted was that predictive modelling helped shift the culture and conversation on their campus toward increased use of data and evidence in decision-making, especially as they relate to both the provision and promotion of services and support for students. The only negative impact noted by respondents was that the modelling identified some students for intervention who otherwise would have been successful (that is, successfully completed the course, retained to second year, graduated or other outcome being predicted). As models are not crystal balls (Finlay, 2014), this negative impact can be mitigated by having more, better quality data to improve the predictive accuracy of the model, but it cannot be completely eliminated.

There was no magic solution for institutions who used predictive modelling, as none of the respondents indicated that they were using a "turnkey" solution. Rather, institutions noted that their model depended on the data that was available to them and the context they operated within. While some institutions changed how they promoted student services, at least one was progressing through a full revision of its academic advising model based on predictive modelling.

It is clear from this research that institutional context is a significant driver of both the focus and the success of modelling efforts. Which measures — overall enrolment levels, year-over-year retention rates or graduation rates, as some examples — were the focus of at least initial predictive modelling efforts was reported in some cases as driven by external factors such as demographic shifts or institutional mandate changes. Other respondents indicated that the primary driver was student success, which is often influenced by factors such as the mix of students served by the institution, the programs and courses offered, and the structure of supports available to students. The successes of predictive modelling efforts were most often linked with improving communication between different areas of the institution and enhancing the institution's ability to make decisions based on evidence.

This research also identified future research directions, such as: how the use of predictive modelling affects student access; how the uptake of predictive modelling interventions is affected by the nature and content of the promotion of the interventions; and followup with the group currently implementing predictive modelling, both in a deeper fashion in the short term for their plans and in several years for implementation lessons.

Table of Contents

List of Figures

List of Tables

Introduction

Canada's population is becoming increasingly well-educated, as evidenced by rising postsecondary enrolment and attainment of postsecondary credentials (Statistics Canada, 2017). However, access, retention and ultimately success for students in underrepresented groups continues to pose a challenge. Identifying students at risk of not succeeding and scaling interventions to provide useful supports to those students is necessary. One approach to addressing this problem is the use of predictive statistical models (Finnie, Fricker, Bozkurt, Poirier & Pavlic, 2017).

The goal of this research, performed on behalf of the Higher Education Quality Council of Ontario (HEQCO) by Plaid Consulting, is to find out whether, where and how predictive modelling is being used effectively to improve student retention in order to be able to inform future predictive modelling projects and new opportunities to leverage institutional data to improve student success. As identified below in the Literature Review, little has been published on predictive modelling for student retention in a Canadian context, a gap this research is intended to address.

As an exploratory study, the primary purpose of this project was to provide an overview of the landscape of predictive modelling in Canada, not a synopsis of technical strategies on how to use predictive models.[3] The focus was on the use of predictive modelling, illuminating whether, where, and how predictive modelling is being used effectively to improve student retention.[4] Through targeted outreach to postsecondary institutions, the departments and individuals most likely to use predictive modelling were identified for participation in a survey. Some responding individuals also opted to participate in followup, in-depth, semi-structured interviews and questionnaires (at their selection) where they could provide more detail on their use, context and challenges as well as any interventions flowing from their predictive models.

The field of predictive modelling is not a new one, nor has its use been limited to a single domain or a particular geographic area. From financial markets and marketing to professional baseball's "moneyball" teams, predictive modelling has found many different applications. Higher education is no different. In this context, predictive models are increasingly popular methods that attempt to predict which students will be "successful" by some measure, such as course completion, retention to second year of studies, meeting a certain aggregate grade threshold or program completion (Campbell, deBlois, & Oblinger, 2007). A more detailed discussion of predictive modelling can be found in Appendix A.

The intent behind predictive modelling is generally to allow for targeted interventions in order to sway the predicted outcomes. This could entail either focusing on the positive individual outcomes and intervening to ensure those outcomes occur or focusing on the negative outcomes and intervening to change those to

3 Some technical insights on the use of predictive modelling are offered in Appendix A and Appendix C.
4 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.

positive outcomes — or a mixture of the two options. Predictive models in and of themselves can neither prescribe outcomes nor interventions — human expertise is required in order to figure out the evaluation and select from the potential actions available in order to use the information from the predictive process to generate the desired future state.

An important point about predictive modelling, particularly in higher education, is that there are few "turnkey" solutions — those which can be used without customizations with existing systems in place. Existing turnkey solutions generally utilize data from a learning management system,[5] which can be done because the concept of successful course completion is relatively standard across higher education. Outside of course completion, the desired outcomes, data available and context (such as curriculum design, student body demographics, institutional characteristics, etc.) generally vary enough from institution to institution and even program to program that any predictive modelling must be customized. Further, all systems need to be reviewed and tuned constantly to account for changes in the outcomes, data or context.

This customization of predictive models in higher education means that to gather information on whether, where and how predictive models are being used and the accompanying challenges and opportunities, information must be sought from many different institutions. Potentially, multiple predictive models may exist at an institution and we need to gather information from multiple departments and individuals within those institutions. Additionally, as predictive models are used globally, there may be information on best practices and lessons learned available from institutions outside Canada.

The distinction between what constitutes more traditional analysis and when that analysis becomes predictive is a grey, blurry one at best. Similar, if not identical, techniques and underlying data are used, with outcome metrics often being similar. One difference may be the timing of the analysis: Is a model being created to explain events which have already happened, or is a future event being predicted?

5 For a detailed description of learning management systems, example systems and the data typically contained in them please see the Glossary in Appendix B.

Research Questions and Methodology

The goal of this research is to find out whether, where and how predictive modelling is being used effectively to improve student retention in order to be able to inform future predictive modelling projects and new opportunities to leverage institutional data to improve student success. The specific research questions were:

a. Which practices and/or principles are used in predictive modelling in the postsecondary context?
b. What are innovative uses of predictive models that influence student retention?
c. What opportunities and challenges are associated with interventions informed by predictive models?
d. How can postsecondary institutions leverage the strengths of predictive modelling to improve student retention? Have the results of predictive modelling changed institutional practices or policies and is there data showing whether these changes have been effective in increasing student access, retention or success?

Ensuring an accurate representation of institutional uses of predictive modelling in student retention requires looking beyond the publicly available literature. Predictive modelling in institutional settings is often undertaken by staff members (see Figure 9, below, for responses to Question 9 that asked participants about who was involved in implementing predictive models at their institution) who are not incentivized to publish in the same way that faculty members are, resulting in minimal information being made publicly available on the modelling undertaken. What literature is available is often out-of-date as models are constantly refined.

In order to get a more complete picture, the project was structured around multiple points of data collection. First, departments and individuals most likely to use predictive modelling were invited to participate in the survey through outreach to postsecondary institutions globally. As predictive modelling is used worldwide, responses were sought from institutions around the world with a particular focus on English-speaking nations where translation would not be required. Second, self-selected survey respondents also participated in followup, in-depth, semi-structured interviews and/or questionnaires (as selected by the respondent) where they could provide more detail on their use, context, challenges and actions flowing from their predictive models.

Phase One: Survey

To reach the widest possible audience with the survey component of the research, we enlisted the aid of postsecondary industry groups. We identified potential audiences through research of industry groups representing recruitment, admissions, academic advising, institutional research and analysis business functions at postsecondary institutions, as those business units were those most likely to utilize predictive modelling as it relates to student retention. Of the postsecondary industry groups identified, three were Canadian-focused, four covered the United States with some international members, three covered other

English-speaking nations (Australia, Ireland, New Zealand and the United Kingdom) and one was global in membership.

Once industry groups were identified, we reached out to the potential groups with background on the survey and information about what we were asking of them and in what timeframe. We initially identified 11 potential groups, of which eight agreed to participate, one declined and two did not respond. Once confirmed, participant groups were sent materials to send to their membership, with a request that they distribute them on the date that the survey opened. A number of Canadian groups utilized their membership email mailing lists to advertise the survey[6] and other groups advertised through means of their membership newsletters and social media groups.[7]

In addition to preliminary recruitment efforts, 315 individuals from Canadian institutions that had not yet responded to the invitation to participate were contacted directly five weeks after the survey was opened. This outreach was completely and purposefully distinct from the initial recruitment strategy. These individuals may or may not have received the initial invitations through the industry groups, and included provosts or vice-provosts, registrars, chief student affairs officers and directors of institutional research. Those contacted depended on the contact information available on each institution's website. Institutional contacts who did not respond were sent a followup email after two weeks and the survey was subsequently closed two weeks after that. The survey was open for a total of 9.5 weeks.

The survey was conducted using the QuestionPro survey platform hosted in Canada (QuestionPro, n.d.). The initial survey was developed with HEQCO and asked representatives of postsecondary institutions about their use of predictive models, including whether predictive models are in use; when and how they were constructed; what interventions are driven by predictive models; the impacts of the models and the associated interventions; and what challenges and successes the institution has found. The complete instrument can be found in Appendix B. To minimize the burden on respondents, only two questions were mandatory: the institution name and whether the institution is using predictive modelling for student retention; additionally, for respondents who indicated they were using predictive modelling, we required a response when we asked whether the respondent would be willing to participate in the second phase of the study.

Overall, 170 responses to the online survey were collected. Of these, 94 were excluded because no was information provided, or because test responses or duplicates of responses were already received. A further six responses were removed from the analysis as they were from international institutions; while the

6 These groups included the Association of Registrars of the Universities and Colleges of Canada (ARUCC, n.d.), the Canadian Association of University and College Student Services (CACUSS, n.d.) and the Canadian Institutional Research and Planning Association (CIRPA, n.d.).
7 These groups included the Australia and New Zealand Student Services Association (ANZSSA, 2015); AMOSSHE, The Student Services Organization in the United Kingdom (AMOSSHE, n.d.); the Association for Institutional Research covering the United States with some international membership (AIR, n.d.); and the American Association of College Registrars and Admissions Officers, also covering the United States with some international membership (AACRAO, n.d.). Each group did an initial advertisement at the August 8 open date as well as a reminder via the same method two weeks later.

invitation to participate was extended to international institutions, the small number of responses made meaningful analysis impossible. This left 70 responses for further analysis — 66 of which were complete, and four of which were partially complete. The survey exceeded the initial target of 40 Canadian responses.

Phase Two: Interview and Questionnaire

During the survey component, respondents were asked to opt in to participating in the interview/questionnaire phase, and whether they would be willing to participate through a telephone-based interview or a questionnaire. All survey respondents were given the choice of either an interview or a questionnaire. The interview requests were sent via email to 12 respondents, while the email questionnaire requests were sent via email to 10 respondents; non-responders in each group received a followup email. Telephone interviews were performed with seven individuals and email questionnaires were sent to and received from three.

The questions were the same in the interview and the questionnaire, but the telephone interview allowed for clarifying questions to be asked while followup questions were not asked of questionnaire respondents. Participants in this phase were asked about the same themes that were present in the survey, but with an eye to receiving additional detail and depth that was not available via the survey. For example, while the survey was structured to provide information on broad data types used in the modelling process and the impetus for undertaking predictive modelling, the interview and questionnaire allowed us to delve into substantially more detail by asking clarification questions. Additionally, the interview and questionnaire allowed us to more easily identify unique practices by asking for additional explanations. The instrument used for both interviews and email questionnaires is available in Appendix E.

Analysis

The responses to open-ended text questions on the survey along with email questionnaire responses and interview notes were loaded into Dedoose (SocioCultural Research Consultants, LLC, 2017) for qualitative analysis. Before upload, data was anonymized and encryption was enabled in Dedoose to ensure the highest levels of protection of confidentiality available. Coding of responses was done by two researchers who coded 100% and 38% respectively of responses to individual questions; Cohen's *kappa* (Cohen, 1960) was 0.77. Noted discrepancies in coding were rectified via discussion.

Literature Review

Background

Predictive modelling, generally, 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). Building on data mining techniques, predictive modelling has been around for decades in various forms and is used across many industries, from financial services to non-profits and government (Finlay, 2014). In the education domain, predictive modelling goes by a number of names, such as educational data mining (Baker & Yacef, 2009), academic analytics (Campbell et al., 2007) and learning analytics (Gašević et al., 2016). The term used in the remainder of this report will be predictive modelling.

The application of predictive modelling in higher education to issues of student outcomes, namely student success and retention, can largely be traced to the adoption of various educational information systems, such as learning management systems (LMS), student information systems (SIS), customer relationship management (CRM) systems, [8] and the broader trend toward increased use of social media and other technology resources. These systems provide institutions with large amounts of data that can be mined for patterns and predictors (Daniel, 2015; Gašević et al., 2016). Predictive modelling within the higher education domain can further be separated into modelling for enrolment management, student success, fundraising or many other outcomes.

Student success and retention have long been studied within higher education research and inform many of the underlying variables used in the various predictive models found in the literature. These models generally rely on some combination of retention factors identified in Tinto (1975, 1987), Bean and Metzner (1985), Astin (1993), or Pascarella and Terenzini (2005). These retention factors may include background factors such as demographic or geographic location information, academic factors such as educational history and attainment, academic and social integration, and external factors such as finances and work or family commitments. More recent research (Kuh, Cruce, Shoup, Kinzie & Gonyea, 2008) has linked student success and student engagement with educationally purposeful activities, such as studying and on-campus activities, with Kahu (2013) viewing retention as a distal consequence of engagement. Finnie et al. (2017a) identified "career clarity" and "education commitment" as potentially significant indicators of a student's likelihood of dropping out of Mohawk College.

Predictive Modelling without Interventions

An early example of student retention predictive modelling is the work of Lam (1984) on predicting drop-out rates of university freshmen at Brandon University using logit regression techniques, and in a similar vein Scalise, Besterfield-Sacre, Shuman, and Wolfe (2000) used logistic regression to identify high-risk, first-term engineering students at the University of Pittsburgh. Other examples of early work include Minaei-Bidgoli and Punch (2003), who worked with genetic algorithms to predict final course grades for students based on LMS interactions, and Morris, Wu and Finnegan's analysis (2005) of high school and standardized test performance data to predict the successful online course completion using predictive discriminant analysis. University of Alabama (UA) graduate students in a 2002 data mining course developed a predictive model that identified 150–200 first-year students each year unlikely to return for their second year; this data was then shared with faculty and advisers for outreach and intervention (Campbell et al., 2007; Davis, Hardin, Bohannon & Oglesby, 2007). The combination of these works provides evidence that predictive models are useful for student retention purposes.

8 For a detailed description and examples of SIS and CRM systems and the data typically contained in them please see the Glossary in Appendix B.

At Purdue University, J.P. Campbell (Campbell, 2007; Campbell et al., 2007) used data from Purdue's LMS to create a model to predict academic success, both for the general population and for a freshman-only model; these models had success rates of 66% and 80% respectively by looking at variables such as SAT or ACT score, overall grade point average (GPA), and composite variables representing LMS usage, assessment, assignments and calendaring. This work was subsequently expanded into the Course Signals project at Purdue.

Unlike other projects, Jia and Maloney (2015) used administrative-only data — that is, avoiding the use of LMS data — from a university in New Zealand to predict both first-year non-completion and second-year non-retention using predictive risk models. While the authors rely more on models used traditionally in risk-management areas such as health care and child protection rather than data mining techniques, the result is a model that looks at many similar factors, such as demography and previous educational experience. Their model was significantly more accurate than using an uninformed model that assumed every course enrolment had the same probability of resulting in non-completion, with the students with the top 10% of risk scores accounting for 29.55% of course non-completions in first year and 23.33% of student non-retention in second year. The authors did not study any interventions related to their model. A key takeaway is that different types of models and analyses lead to similarly useful results when predictive modelling is applied to student retention.

Interventions

Arnold and Pistilli (2012) provide information on the evolution of Course Signals (CS) post-2007. CS was implemented in Purdue's LMS, where instructors were able to run the predictive model for their students providing instructors with a "traffic signal" indicator. This signal, in turn, was placed on the student's LMS course homepage, with a green signal for those with strong likelihood of success, yellow for those who may have issues succeeding and red for those who are likely to struggle. Faculty would then decide on specific interventions with students, including email messages, text messages, referrals to academic advisers and one-on-one meetings with the instructor. Students who had at least one course that utilized CS were retained at significantly higher rates than students who did not have a course utilizing CS, and students with two or more courses with CS were retained at even higher rates. Additionally, student retention rates improved when CS courses occurred earlier in the student's career. One outcome of this study is the finding that course-level interventions can have a positive impact on program- and university-level retention.

Also working primarily with LMS data, the Open University (OU) — a distance-learning institution in the United Kingdom — has been building a predictive model to identify students at risk of not successfully completing a course, beginning initially with two introductory courses but having expanded to 18 courses as of 2015 (Kuzilek, Hlosta, Herrmannova, Zdrahal, & Wolff, 2015). The OU's project involved several different prediction models and utilized both demographic and LMS data, and ultimately intervening with students to try to bring them back on track; interventions are generally in the form of an outreach phone call from specialized student support teams, but Kuzilek et al. do not provide information on the effectiveness of the interventions. The primary finding was the criticality of the early identification of students at risk of not successfully completing a course.

While other predictive models in student retention were focused on a single institution, several later projects looked at creating models that utilized data from multiple institutions. In one example, Jayaprakash, Moody, Lauría, Regan, and Baron (2014) reported on their work on the Open Academic Analytics Initiative (OAAI). The OAAI used LMS and administrative performance data from Marist College in New York State to build a predictive model for student success that was then tested at several community colleges and historically black colleges and universities. The model was used to inform interventions for those students predicted to be at risk of not being in good standing academically. OAAI's work is based on open source software and is — at least in theory — transferable to other academic contexts.

The interventions studied by Jayaprakash et al. (2014) include "awareness messaging" based on and very similar to the email interventions used by Course Signals at Purdue and the creation of an "online academic support environment" (OASE) within the institution's LMS that provided open educational resources for students to use. The OASE resources promoted awareness of support mechanisms, self-assessment tools, and scaffolding to improve study habits and refresh fundamental content. The study, run across all participating institutions in selected (but not identified) courses in spring and fall 2012 using a model created and tested at Marist College in fall 2010 and spring 2011, separated students identified as academically at risk into a control group, an "awareness messaging" group and a group provided with access to the OASE content. Academically at-risk students receiving interventions generally performed better than the control group (receiving grades 6 percentage points higher on average), however the intervention groups generally showed higher course withdrawal rates than the control group. Early withdrawal may in fact be a positive outcome as, if students feel they would not be able to improve their grade enough to pass, withdrawal may allow them to avoid a negative impact on their transcript. This study suggests that models and interventions can be used across contexts while retaining much of their power.

Another multiple institution initiative was conducted at the University of Maryland University College (UMUC), which undertook a four-year project determining factors that could predict student success following transfer from two community colleges in the Maryland system (University of Maryland University College, 2015). They used observations of demographic factors, community college course-taking patterns and performance, online course engagement, and early career UMUC performance to predict future performance, re-enrolment (enrolment in their second term), retention (enrolment within a 12-month window following the first term) and graduation at UMUC. UMUC found that different factors were predictive of different outcomes — for example, gender was predictive for first-term performance, re-enrolment and retention but not graduation, while whether a student took a math course at the community college prior to transferring to UMUC was predictive of graduation but not first-term performance, re-enrolment or retention. UMUC additionally studied four different interventions performed on students after their transfer from community college to UMUC. In three of the interventions — use of a student resource checklist, mentoring and a pre-enrolment "Jumpstart" onboarding course — did not show statistically significant differences in term GPA, successful course completion, or re-enrolment between control and experimental groups, while the final intervention using in-person tutoring did show significant differences. Of note is the small size of the experimental groups, so these results should be treated with caution.

Predictive Modelling for Student Retention in Canada

In the Canadian context, Finnie et al. (2017a) report on work at Mohawk College, where the research and implementation of a predictive modelling system began in 2012. Using data from students who entered Mohawk between 2005 and 2012, the project created predictive models for students who entered in fall 2013 and fall 2014 and categorized students into low-, medium- and high-risk student risk classifications (SRCs) that represented their likelihood of retention in both their second term and their second year of studies. The variables in the model included information on the student and the student’s program, responses to an entrance survey and scores on reading, writing and mathematics assessments conducted prior to the student’s first term. The results showed that 9% of students in the 2013 and 2014 entry classes in the low-risk SRC left Mohawk, with 13% of medium- and 23% of high-risk students leaving; the model ended up being fairly accurate in the predictions. While at this stage the predictive model was not coupled with interventions for students, the authors note that as student risk level (as determined by the model) went up, students were more likely to seek out advising resources at least once in their first year at Mohawk.

In the second phase of Mohawk’s project, students in the fall 2015 entering class were randomly assigned to one of three groups — a control group that received an email before classes began advertising advising services, an experimental group that received up to three additional emails and a phone call encouraging participation in a group advising session, and a similar experimental group that utilized one-on-one advising. While the group assignment was not informed by the predictive model from Finnie et al. (2017a), Finnie et al. (2017b) report on the effects of the treatment on the experimental groups by the predictive model’s SRCs. The results are not particularly clear, though show that for some groups there were statistically significant improvements in leaving rates. High-risk students in group advising were 4.8 percentage points less likely to leave ($p=0.1$) after their first term, and low-risk students in group advising were 4.9 percentage points less likely to leave ($p=0.1$) after their first year, with males in that group being 9.8 percentage points less likely to leave ($p=0.05$).

In summary, the previous work in predictive modelling for student retention spans geographic boundaries, with work in North America, Europe and Oceania. While there has been a lot of focus on LMS data as a major data source, other models are focusing either on using only administrative data or directly gathering information from students. Few of the identified research projects have been turned into ongoing operational systems, with proprietary systems instead having moved into the space; notable exceptions are at the University of Alabama where the described system was used for at least five years, Purdue’s Course Signals project and The Open University’s system.

Within the literature available on predictive modelling for student retention, very little focuses on the Canadian context. Lam's (1984) work utilized student data from Brandon University, while Finnie et al. (2017a, 2017b) reported on work at Mohawk College. The lack of literature that involved Canadian institutions was a key impetus for this project.

Results

The survey included a landing page, where potential respondents were provided with information on the reasons the project was being undertaken, a link to the privacy policy that governed the collection of the information, and contact information for both Plaid Consulting and HEQCO should the respondents have any questions. When the "Next" button on this page was clicked, the survey software logged a survey response; in total, 170 responses were logged.

Of these 170 responses, we excluded 100 from our pool of responses to review and analyze: in 58 cases the Next button was clicked but no information was provided; in six cases, responses were from outside Canada; in five cases the respondent entered meaningless data, such as an institution name of "test" or "ghjg"; in 15 cases the respondent entered an institution name but did not provide an answer to Question 4 asking if the institution used predictive modelling for student retention; in four cases the respondent answered Question 4 with a "yes" but did not provide an answer to Question 5 asking how they used predictive modelling; and in 12 cases the response was a duplicate to a response we did include in the analysis. This left us with 70 usable responses for analysis, of which 66 were complete and four were partially complete. Completeness indicates that respondents pressed "Submit" on the final page of the survey, and as the survey was linear with no method to jump over questions we know every question was viewed, though not necessarily answered. The partially complete responses provide information up to, respectively, Questions 9, 18, 19 and 21 of the survey. The partially complete responses were included in the analysis as a number of other responses, though complete, did not answer some of the survey questions, and we saw no reason to exclude the partially complete group simply because their unanswered questions came at the end of the survey.

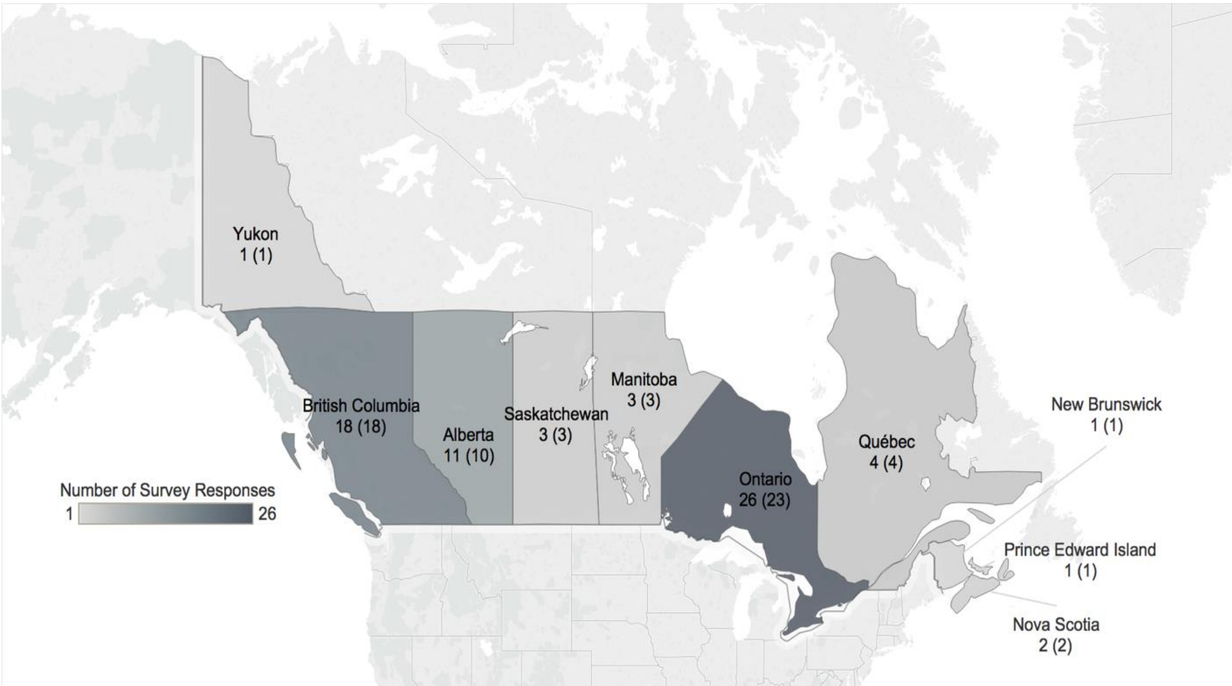
In some cases, multiple responses were received from different areas of the same institution, and in total 70 responses were received from 66 Canadian institutions; the 12 duplicate responses removed were those cases where clearly the responses were from the same individual in the same area of the institution. Responses were received from institutions in nine of 10 provinces (with the exception being Newfoundland and Labrador) and the Yukon was the only one of the nation's three territories to be represented. Ontario saw the highest number of responses with 26, followed by British Columbia (18) and Alberta (11). Table 1, below, shows the number of respondent institutions by Canadian province or territory and an approximate response rate for each. Note that this table counts institutions rather than individual responses, while the total number of institutions in each province or territory includes publicly funded postsecondary institutions, and institutions that are members of Colleges and Institutes Canada, Polytechnics Canada or Universities Canada.

Table 1: Survey Response Rates by Province/Territory

Province/Territory	Number of Respondent Institutions	Total Number of Institutions	Response Rate
Alberta	10	20	50.0%
British Columbia	18	27	66.7%
Manitoba	3	8	37.5%
New Brunswick	1	7	14.3%
Newfoundland and Labrador	0	4	0.0%
Northwest Territories	0	2	0.0%
Nova Scotia	2	10	20.0%
Nunavut	0	1	0.0%
Ontario	23	54	42.6%
Prince Edward Island	1	3	33.3%
Québec	4	68	5.9%
Saskatchewan	3	15	20.0%
Yukon	1	1	100.0%
Grand Total	66	220	29.1%

See Figure 1 for a map of the responses by province or territory. In some cases, multiple responses were received from different parts of a single institution, and the first number shown is the number of survey responses received, while the second number in parentheses is the number of distinct institutions that responded.

Figure 1: Map of Responses to the Predictive Modelling for Student Retention Survey



Note: A darker colour indicates more responses; numbers represent number of survey responses and those in parentheses indicate the number of distinct institutions responding.

Most of the survey responses come from individuals with either oversight responsibility or a role in developing the predictive model. Figure 28 in Appendix C provides the detailed breakdown.
At the end of the survey, the 25 respondents from institutions currently using predictive modelling were asked if they would be willing to participate in either a followup interview via telephone or a questionnaire over email. Seven interviews and two email questionnaires were subsequently completed by participants in phase two of the data gathering.

Figure 2 maps the participants from this phase; unlike phase one, in phase two each institution provided only a single response and the colouring and numbering represents both the number of responses and the number of institutions. As the questions for participants during the phone interviews were substantially similar to those provided to participants in the email questionnaires, the analysis was performed on the responses to both as a whole. The initial questions for both methods were the same, though the interview allowed for additional questions to be asked in order to clarify or expand on certain points.

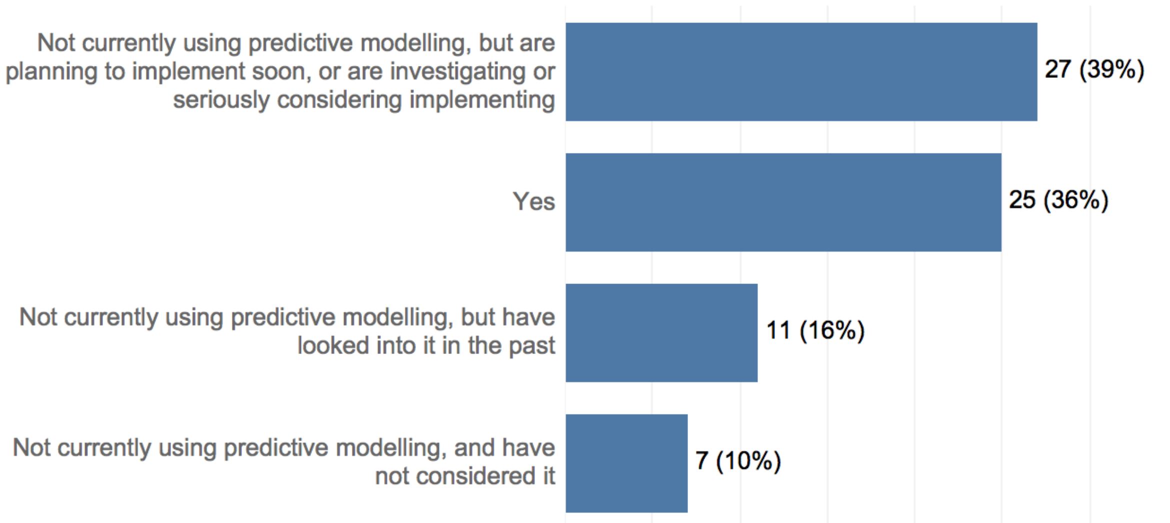
Figure 2: Map of Participants in the Predictive Modelling for Student Retention Interview and Email Questionnaire Phase

Use of Predictive Modelling

The survey results have shown that the use of predictive modelling at postsecondary institutions for student retention purposes is increasing, particularly since 2013. Most institutions using, or considering using, predictive modeling note that the impetus for doing so was to support student success, with smaller numbers citing institutional requirements or priorities. Information gathered in interviews and/or questionnaires showed that in almost all cases, predictive modelling is being used across the institution rather than within a specific department or program, and the most common uses are for enrolment planning and admissions projections. Other reasons cited for implementing predictive modelling include enhancing student success and identifying students at risk of not being academically successful. The impetus for using predictive modelling typically comes from senior ranks within the institution or offices responsible for enrolment planning/services.

The final response analysis group consisted of 70 survey responses. Of these, 25 responses (36%) indicated that their institutions are currently using predictive modelling for student retention purposes and a further 27 (39%) that they were investigating, seriously considering or planning to use it. These results are shown in more detail in Figure 3 below.

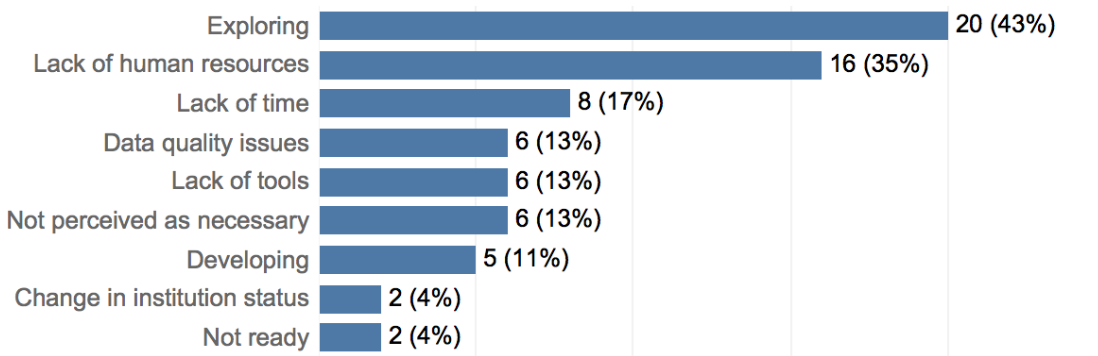
Figure 3: Responses to Question 4, "Is your institution currently using predictive modelling for student retention purposes?" (Select One)



Respondents who indicated any response other than "yes" on Question 4 were given the opportunity to elaborate on why they were not currently using predictive modelling, and whether they thought that might change in the future. These responses were qualitatively coded and the results can be seen in Figure 4. Of these 45 respondents, 20 (43%) indicated that they were currently exploring using predictive modelling but were not yet at the point of definitely implementing a system. A number of responses indicated resourcing was an issue, either in terms of people (16, 35%), time (eight, 17%), or tools such as appropriate software and sufficiently powerful hardware (six, 13%), while six respondents (13%) identified issues with data quality or understanding as holding them back.

An additional concern, cited twice (4%), was a change in institutional mandate as the resulting programming changes mean that historical retention may not be predictive of future retention, and these institutions may have to wait for several years before predictive modelling can be reasonably pursued. Six respondents (13%) indicated that predictive modelling was not perceived as a need in their institution, with three of these responses citing already high retention and graduation rates as the reason it had not been pursued; another two (4%) indicated the institution was not ready for predictive modelling, without indicating why. Finally, five respondents (11%) indicated that predictive modelling would be moving forward and they were currently in a development phase, with one indicating a fall 2017 pilot and another a fall 2018 launch.

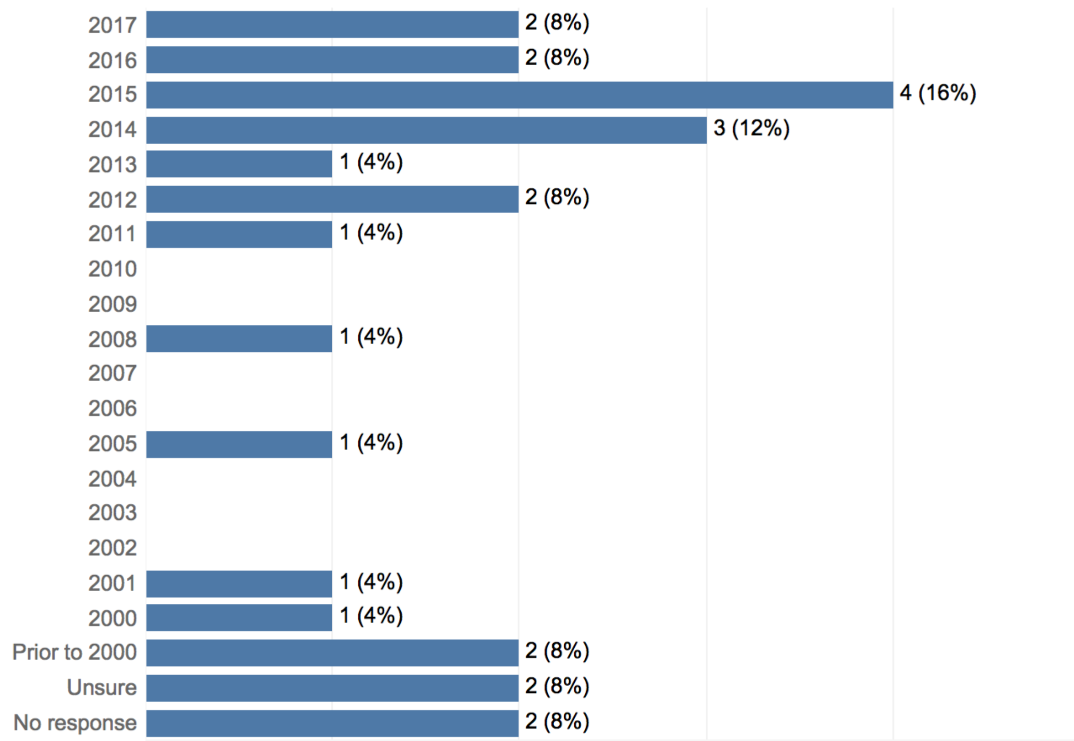
Figure 4: Qualitative Coding of non-"Yes" Responses to Question 4, "Is your institution currently using predictive modelling for student retention purposes?"



The 25 respondents who indicated in Question 4 that their institution was using predictive modelling were directed to the remainder of the survey, while the other respondents were sent directly to a page thanking them for their participation in the survey. The following analysis focuses on the 25 respondents who were using predictive modelling.

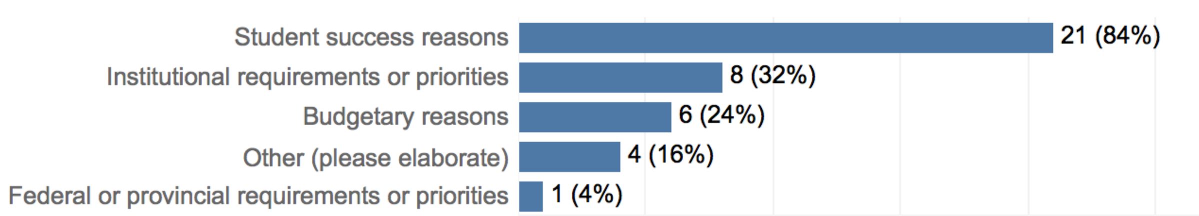
Use of predictive modelling is a relatively recent phenomenon at postsecondary institutions with many of the responses showing that predictive modelling has only been in place since 2013 or later, with 2014 through 2017 garnering 11 responses (39%). Prior to 2013, adoption of predictive modelling occurred at a much slower pace (see Figure 5).

Figure 5: Responses to Question 7, "When did you begin using predictive modelling for student retention purposes?" (Select One)



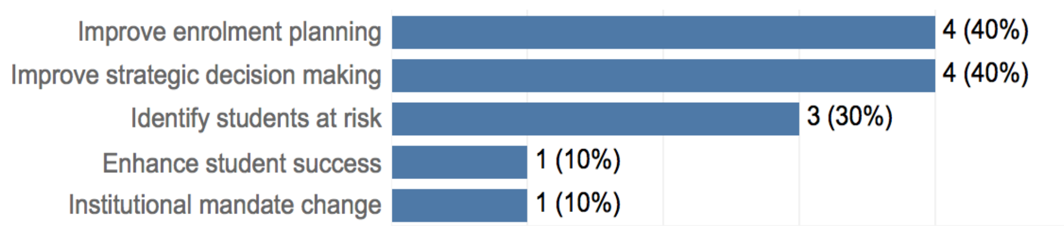
When asked why they began using predictive modelling, most institutions (21, 84%) indicated they began for student success reasons, with smaller numbers citing institutional requirements or priorities (8, 32%), budgetary reasons (six, 24%), or federal or provincial requirements or priorities (one, 4%). Among those that responded "other," one indicated enrolment planning, one academic improvement, one tailoring student success interventions and one offering a better curriculum and understanding future student needs. (See Figure 6.)

Figure 6: Responses to Question 14, "What was the impetus for your use of predictive modelling for student retention?" (Select All That Apply)



These issues were explored in more detail with the nine institutions that either participated in an interview or completed a questionnaire (see Figure 7). The most cited reasons for why participants began using predictive modelling included improving enrolment planning (four, 40%), improving strategic decision-making (four, 40%) and identifying students at risk of not being academically successful (three, 30%). Enhancing student success was identified by one participant (10%). An institutional mandate change was identified by one institution (10%), which relayed that the mandate change and associated changes to programs made the historical "ruler method" (where many years of historical data were used to ballpark future years) invalid and that new techniques were required to interpret data on much shorter timeframes.

Figure 7: Phase Two Participants' Rationale for the Adoption of Predictive Modelling



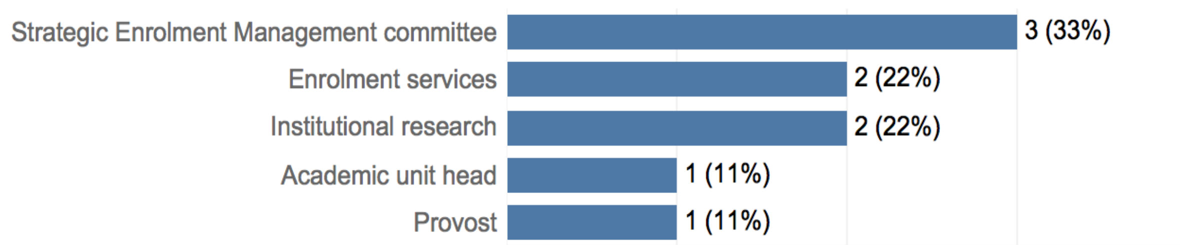
Note: Some responses received multiple codes.

Almost all survey and interview participants (eight of nine, 89%) reported using predictive modelling across the institution rather than within a specific department or program.

Predictive modelling was performed on both ad hoc and regular bases depending on the purpose of the model and the availability of data. Participants reported having models that were both aggregate — that is, predicted retention for a group of students without predicting retention for any particular students — and individual. These results are shown in more detail in Appendix C (see Figure 37).

The request to begin using predictive modelling most often came from either senior ranks of the institution, such as a strategic enrolment management (SEM) committee (three, 33%) or the provost's office (one, 11%), from offices responsible for planning such as within the institutional research office (two, 22%), or from enrolment services (two, 22%). In one case (11%) the request originated from the academic unit head and this was the case where the modelling was only done for a particular program rather than more broadly across the institution (See Figure 8).

Figure 8: Phase Two Participants' Indication of Where Predictive Modelling Request Originated

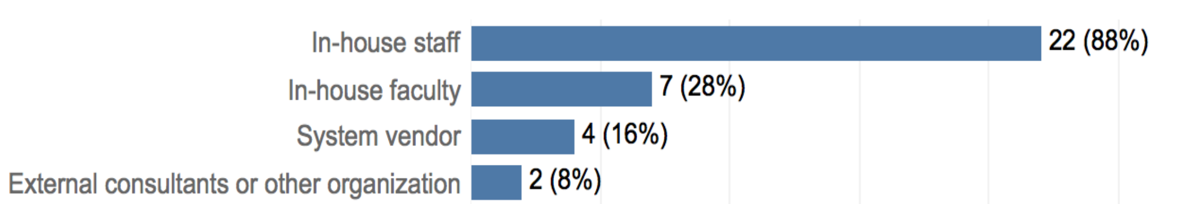


Institutional Stakeholders

Most predictive modelling systems currently in use at postsecondary institutions were implemented by in-house staff and/or faculty. In some cases the system vendor or an external consultant or other organization was used. The results of the predictive modelling were made available primarily to administrative managers, academic advisers, other administrative staff, unit heads, SEM committees and other senior administrators. Few respondents indicated that faculty/non-faculty instructors were given access. In two cases the results were made available directly to the student. In most cases, the actual predictions from the system were made available through custom reporting.

Of those institutions currently using predictive modelling, 22 of 25 survey respondents (88%) indicated that their systems were implemented by in-house staff. In some cases in-house faculty were also involved (seven respondents, 28%). In four cases (16%), the system vendor was used; and, in two cases (8%), an external consultant or other organization was used (Figure 9). Of those that selected more than one of the response options, six (24%) indicated that the system was put in place by in-house staff, in-house faculty and the system vendor. A further three (12%) indicated that the system was put in place by in-house staff and the system vendor. No other combination garnered more than one response, and one respondent provided no response to this question. When "external consultants" was selected, respondents were given the option of providing the consultant's organization, with two indicating Noel-Levitz.

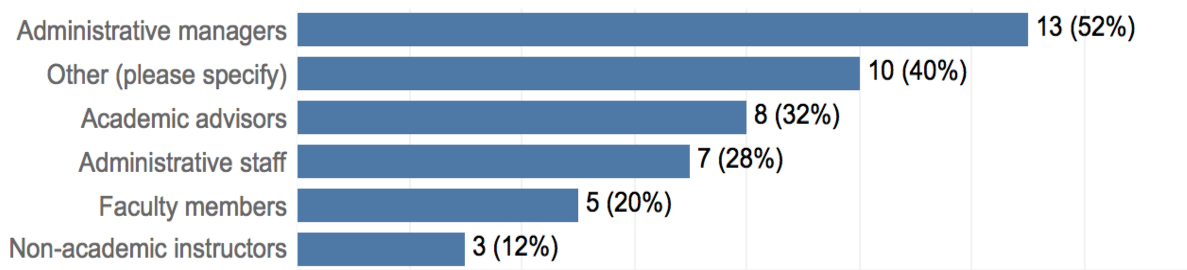
Figure 9: Responses to Question 9, "When you originally implemented your predictive modelling for student retention, who was involved in the implementation?" (Select All That Apply)



When asked who has access to the results of the predictive modelling, the survey showed that the majority of systems provide access to some administrative managers (13, 52%), while in eight cases (32%) access is provided to academic advisers and in another seven (28%) to other administrative staff. Few systems provide access to faculty members (five, 20%) or non-faculty instructors (three, 12%).

The 10 (40%) responses in the "other" category included: academic unit heads (two responses), strategic enrolment management committee (two), senior administrators (two), direct to the student (two) and on a case-by-case basis (three). These results are shown in Figure 10.

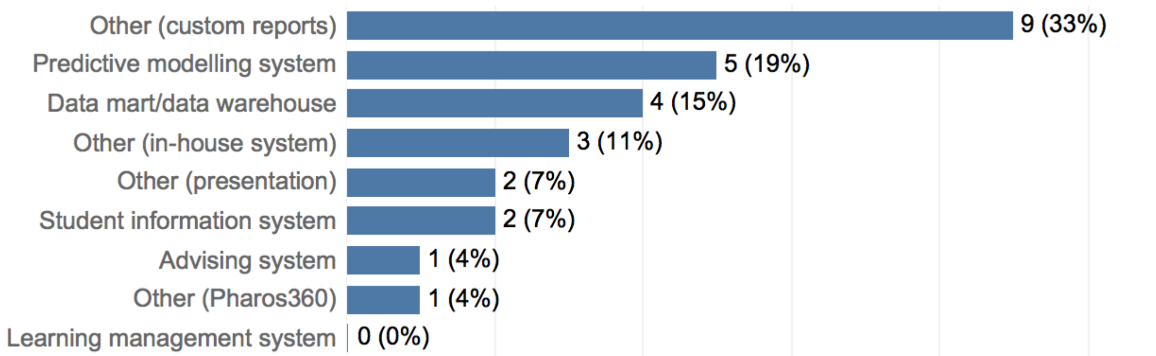
Figure 10: Responses to Question 16, "Who has access to the predictions from the predictive modelling system?" (Select All That Apply)



When asked how they maintained security around the predictions ensuring only authorized individuals could access the information, interview/questionnaire participants reported that, in general, access to information where the predictions were done in aggregate form was less tightly restricted than if the predictions were done at the individual level. In two cases (20%) the predictions were uploaded onto an operational system — in one case to a CRM system, and in another to the SIS — where appropriate security was applied. In other cases with individual data, the information was stored only within a database with extremely limited access. Where the predictions are aggregate, the information was more often made available to groups across the institution, though ensuring that the information did not make it outside the institution was required.

Survey results showed that, in most cases, the actual predictions from the system are made available either through the system itself (five, 29%) or through custom reporting (nine, 33%) as shown in Figure 11. A smaller number of responses were seen for data mart or data warehouse (four, 15%), student information system (two, 7%) and advising system (one, 4%); the LMS option was not selected at all. Among the "other" group, three respondents (11%) indicated another system, two respondents (7%) indicated predictions are provided in presentations, and in one case (4%) through the Pharos360 system previously mentioned.

Figure 11: Responses to Question 17, "How are the predictions made available to those with access?" (Select All That Apply)



Note: Total responses = 30; "Other" responses separated by qualitative coding.

How is Predictive Modelling Being Used?

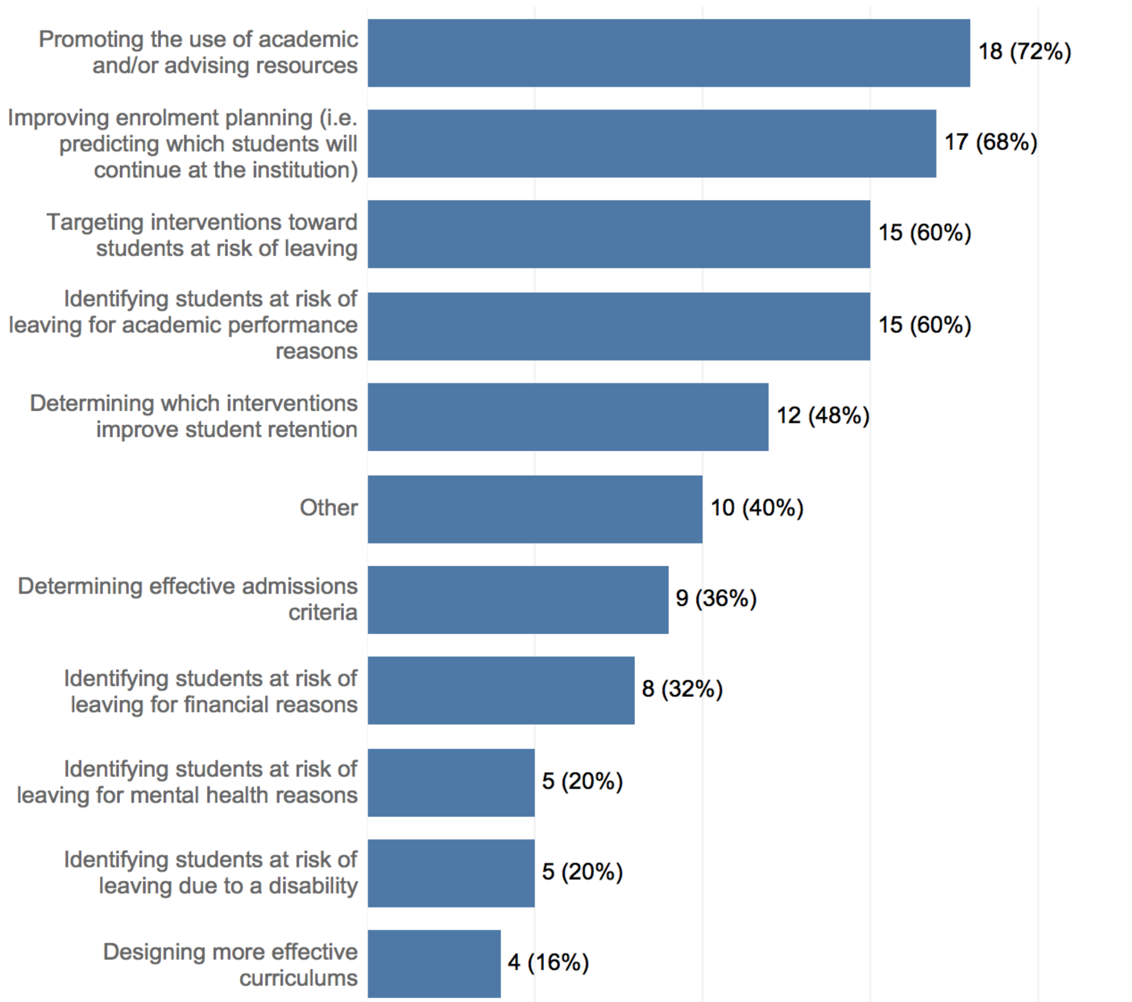
Predictive models are used for many reasons associated with enhancing student success including identifying vulnerable students, targeting interventions, promoting student supports and improving enrolment planning. The primary target group for models is first-year undergraduate students. Perceptions of respondents vary with regard to the accuracy of predictions for different student groups. The data used in predictive models is most often gathered from the institution's SIS. Data may also be used from other systems such as financial aid, advising, student engagement and learning management systems. The types of information commonly used in predictive models include student demographics, location, previous educational history, student surveys, standardized tests, and admission and application information. More than half of the respondents indicated that they use predictive models to inform specific student retention intervention such as the promotion of available support services, advising, mentoring or self-assessment tools.

Types of Uses

In Question 5, respondents were asked how predictive modelling was being used at their institution, and the results can be seen in Figure 12. This was presented as a select-all-that-apply question, and options that garnered a majority of the 25 possible responses include "identifying students at risk of leaving for academic performance reasons" with 15 (60%), "targeting interventions toward students at risk of leaving" with 15 (60%), "promoting the use of academic and/or advising resources" with 18 (72%), "determining which interventions improve student retention" with 12 (48%) and "improving enrolment planning" with 17 (68%). There is substantial overlap in the nature of these uses — targeting interventions and determining the effectiveness of those interventions go hand-in-hand, for example.

Fewer respondents indicated that they used predictive modelling to identify students at risk of leaving for non-academic reasons, such as mental health (five, 20%), disability (five, 20%) or financial (eight, 32%) reasons, determining effective admissions criteria (nine, 36%) or designing more effective curriculums (four, 16%). In the "Other" category one (4%) respondent indicated use of predictive models for each of course scheduling, housing, assessing the connection between secondary school attended and postsecondary performance, and the setting of retention targets, while three (11%) respondents indicated use in enrolment planning but for non-retention purposes — one respondent indicated looking at graduation rather than retention rates, another using a budget rather than a retention lens, and the last course offerings — and three (11%) indicated looking for at-risk groups without identifying particular students within those groups.

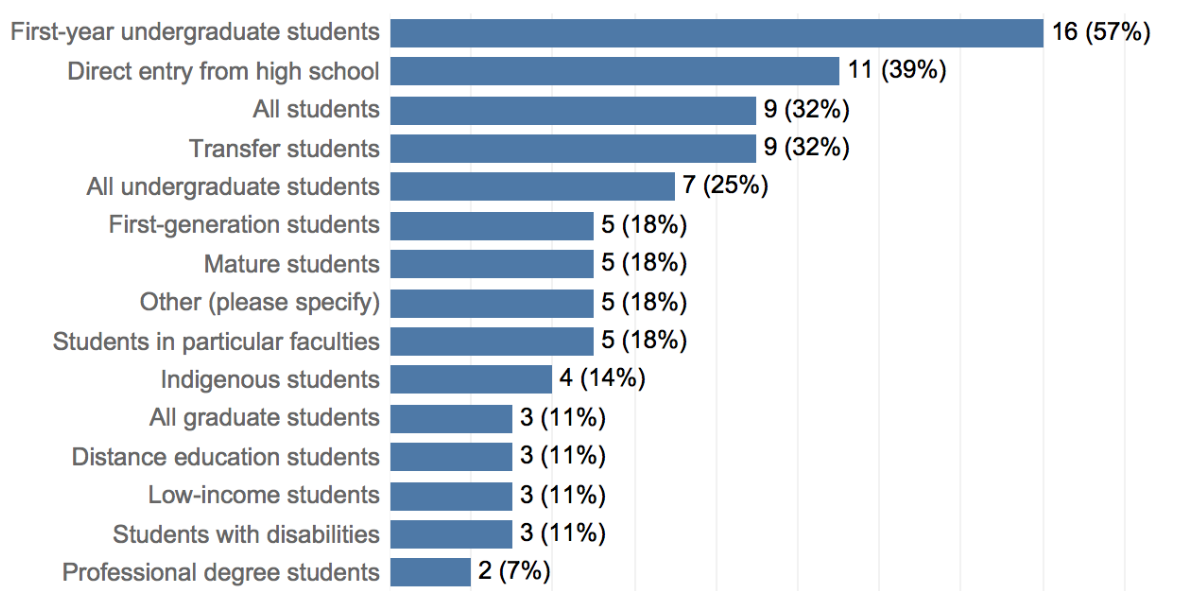
Figure 12: Responses to Question 5, "You indicated that your predictive model is used for student retention purposes. In what ways does your institution use predictive modelling?" (Select All That Apply)



Student Populations

Most predictive models are used for first-year undergraduate students, with 16 (57%) of respondents selecting this answer. As shown in Figure 13, this was the only response that a majority of respondents chose. Other responses frequently chosen include students entering directly from high school (11, 39%), transfer students (nine, 32%), all students (nine, 32%) and all undergraduates (seven, 25%). Of the nine respondents that selected "all students," five also selected at least one additional category, with one choosing all options except "other"; the same is true for "all undergraduate students" with most (nine of 10) respondents also selecting another category. In the "other" category, two respondents indicated that they use predictive modelling for students in residence, with other respondents mentioning scholarship students (one respondent), first-year master's students (one respondent) and students in particular faculties. When asked to elaborate on "students in particular faculties" two of the five respondents indicated they used predictive modelling for all faculties/programs but included faculty or program as an input model to the variable. Other respondents indicated that they used predictive modelling for direct entry programs (one respondent) and non-cohort programs (one respondent), and one respondent specifically mentioned faculties of law and engineering.

Figure 13: Responses to Question 13, "For which student populations does your institution use predictive models for student retention?" (Select All That Apply)



When asked about how accurate their predictive modelling was, in terms of the percentage of students accurately predicted, in half of the cases (36 of 72) the respondent was not sure about the system's accuracy; of those who did respond in some way, 27 of 72 (38%) of respondents indicated 70% accuracy or higher. See Table 2 for further details and a breakdown by student group.

Table 2: Responses to Question 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 responses
First-year undergraduate students	1	1			2	2	8	**14**
Direct entry from high school	1	1	1		2	1	3	**9**
All students				4		2	2	**8**
Transfer students	1				2	1	3	**7**
All undergraduate students	1			1	1	3		**6**
Mature students				1	1		3	**5**
Indigenous students		1			1		2	**4**
First-generation students				1			3	**4**
Other (as identified in Q13)				1			3	**4**
Students in particular faculties							3	**3**
Low-income students							2	**2**
Distance education students					1		1	**2**
Students with disabilities		1					1	**2**
All graduate students							1	**1**
Professional degree students							1	**1**
All responses	**4**	**4**	**1**	**8**	**10**	**9**	**36**	**72**

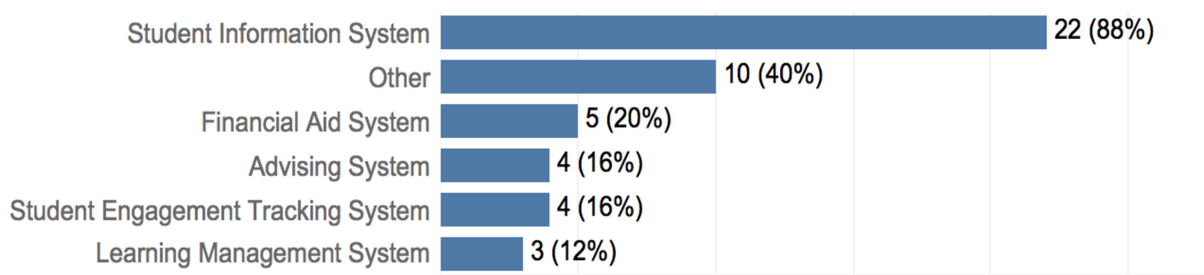
Data Used

The most common source of information for predictive modelling systems is the SIS, with 22 respondents (88%) indicating that their system used some information from the SIS (see Figure 14). Of the remaining specific options, all were used by between three (12%) and five (20%) respondents. Systems in the "other" category included survey and questionnaire systems (three), prospective student customer relationship management systems (two), an early alert system (one), data that is held outside of any particular system, such as in Excel spreadsheets (one), an unnamed SIS (one; not reflected in the 22 indicating SIS), and one respondent that provided no further elaboration. One respondent did not provide any response.

The most common combination of responses was student information system and other (five, 20%); no other combinations garnered more than one response. Details of the combinations of responses can be seen in Figure 15.

Among the SIS systems represented, Ellucian's Banner had nine (41%) responses and Oracle's PeopleSoft Campus Solutions had three (14%), while two respondents (9%) indicated an in-house built ("home-grown") system. Other responses included "Ellucian" (which could be either Banner or Ellucian's Colleague product), "Oracle" (which could be PeopleSoft or the Oracle Student System), while Skytech and CrossRoad were each mentioned once. One respondent did not know, while three did not provide any information on the product. Learning management systems mentioned include Blackboard, Desire2Learn and Moodle; advising systems were End2End and two home-grown systems; financial aid system was Navison (others did not know or pulled financial aid data from their SIS); and student engagement systems included ezRecruit, CampusLabs and home-grown systems.

Figure 14: Responses to Question 11, "What systems do your predictive models for student retention gather information from?" (Select All That Apply)



The majority of respondents indicated that they used information on student demographics (20 respondents, 80%), location (14 respondents, 56%), and previous educational history either at their institution (17 respondents, 68%) or in secondary school (16 respondents, 64%) in their predictive models (see Figure 16). LMS interactions was low in the number of responses (three respondents, 12%), which matches the LMS responses from Question 11. Self-assessment questionnaires (eight respondents, 32%) tended to be locally developed (four of eight responses, 50% of self-assessment questionnaires), though both the Canadian University Survey Consortium (two of eight responses, 25% of self-assessment questionnaires) and the National Survey of Student Engagement (two of eight responses, 25% of self-assessment questionnaires) were mentioned as outside surveys used. Among standardized tests (four, 16%), the SAT and ACT were included in one response, while two responses included English language tests, one response included English and mathematics placement exams developed locally and one response indicated the law school admission test (LSAT). Other information (eight, 32%) included admission and application information, such as when a student applied and to which programs (four of eight responses, 50% of respondents indicating other information); employment status (one of eight responses, 13% of respondents indicating other information), financial information, including financial aid (four of eight responses, 50% of respondents indicating other information); on-campus residence status (one of eight responses, 13% of respondents indicating other information); and student/faculty reviews (one of eight responses, 13% of respondents indicating other information).

Figure 15: Combinations of Responses to Question 11, "What systems do your predictive models for student retention gather information from?"

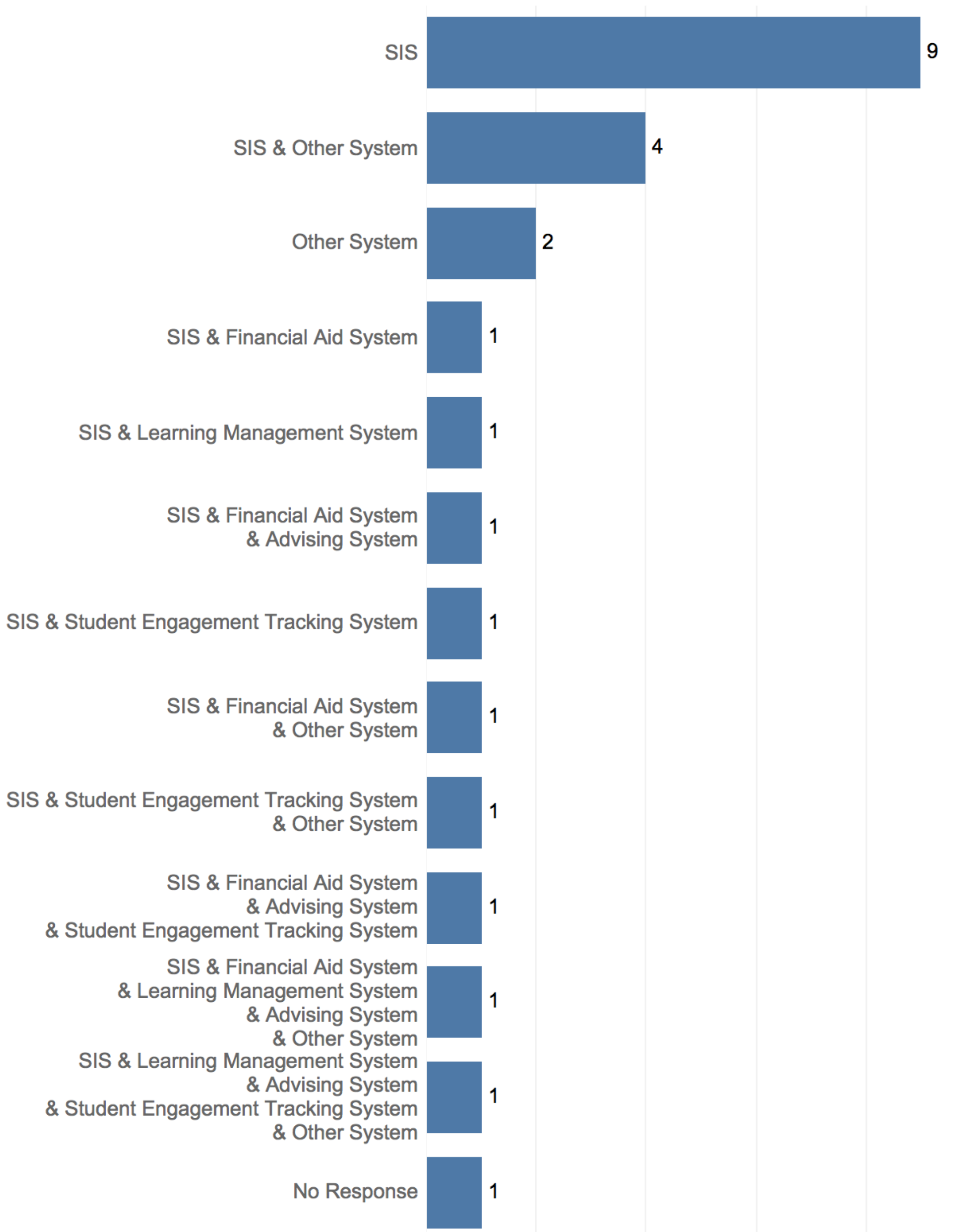
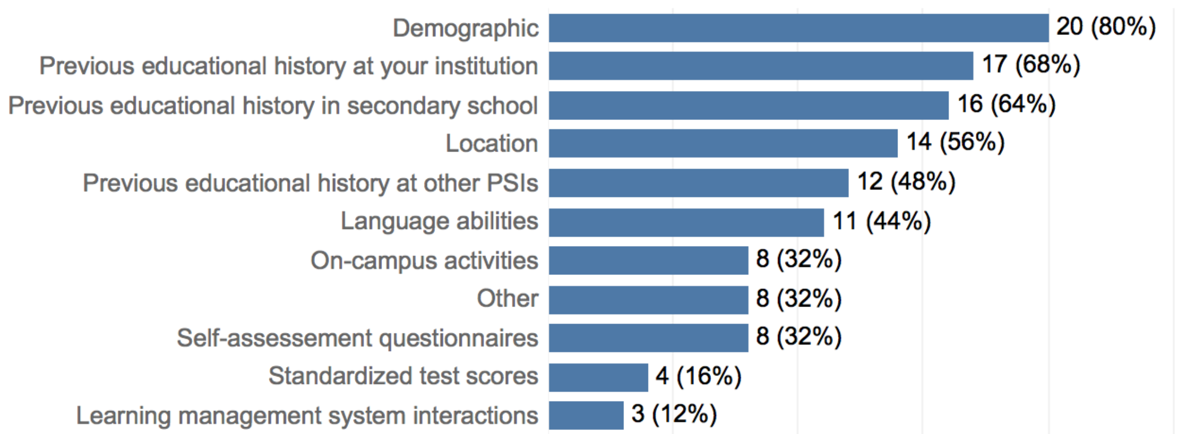


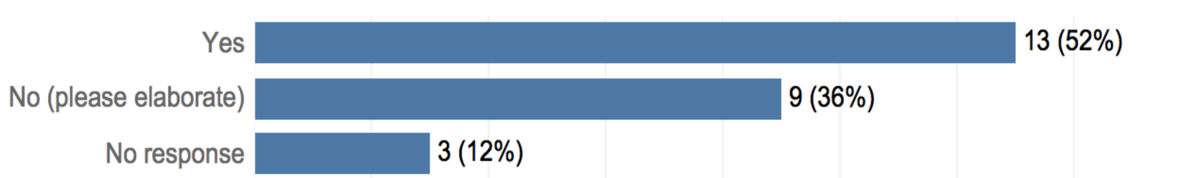
Figure 16: Responses to Question 12, "What types of information do your predictive models use for student retention?" (Select All That Apply)



Interventions

When asked if they use their predictive modelling to inform specific student retention interventions, 13 (52%) respondents indicated they did, while nine (36%) respondents indicated they did not; three respondents (12%) did not answer the question (See Figure 17). Among those in the "no" group, three indicated that they hoped or planned to at some point but that their models were still in the process of being tested and refined, while two respondents indicated the models were not used for those purposes, two other respondents indicated that the information was provided to the students to determine how to proceed, and one respondent indicated they had provided examples and suggestions to areas that would run the interventions but had not yet seen any uptake.

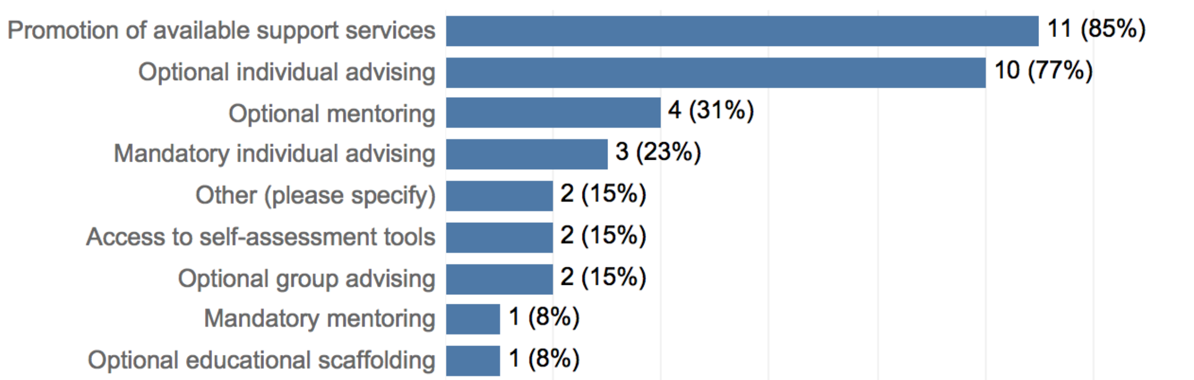
Figure 17: Responses to Question 20, "Do you currently use predictive modelling to inform specific student retention interventions?" (Select One)



Part of the intention of Question 20 was that respondents would only be asked Questions 21 through 25 if they responded "yes." However, this logic was mistakenly not implemented in the final survey. The results below for Questions 21 through 25 are only provided for the 13 respondents who indicated that they do use predictive modelling to inform specific interventions.

Question 21 asked which interventions were used alongside predictive modelling (see Figure 18). The largest groups of responses were the promotion of available support services (11, 85% of 13 respondents who answered "yes" to Question 20) and optional individual advising (10, 77%). The remaining options have noticeably fewer responses, with only a small number of positive responses to optional mentoring (four, 31%), mandatory individual advising (three, 23%), optional group advising (two, 15%), optional educational scaffolding (one, 8%), access to self-assessment tools (two, 15%), other (two, 15%) and mandatory mentoring (one, 8%). Among those responding "other," one respondent indicated they have a peer tutoring program, while one other respondent indicated they make programming and advertising decisions based on predictive modelling at an aggregate level but not at the level of a specific student. One additional response to "other" was "early alert" but no further information on the specific intervention associated with the early alert was provided.

Figure 18: Responses to Question 21, "Please indicate the different types of interventions your institution uses based on your predictive models." (Select All That Apply)



Many interview/questionnaire participants reported new interventions related to predictive modelling. Some interventions were in place during the admissions process, where the type and content of communications with applicants shifted due to information from the models, while another participant reported changes in the sources of applicants. Other interventions were used with students while they attend the institution, such as improved outreach to students, while one participant reported a wholesale shift in the institution's advising model. Participants also noted that not all interventions are one-on-one, with one participant specifically noting the use of predictive models to convince senior leadership to increase the number of course sections offered; this increase led to improved retention rates as students could get more of the classes they wanted.

Participants were also asked about what helped them shift predictive modelling from a more theoretical tool to an applied context. While one participant noted technical skills in data analysis and data wrangling, most responses centred on having high-level support and being able to generate some early wins using predictive models. Having early wins was noted as leading to stakeholders having confidence in the decisions made based on the predictive models.

Opportunities and Challenges in Predictive Modelling

As predictive models have not been in use for many years in most institutions, respondents report that it is often too early to tell what the impact is. Though in some cases institutions report that they see a positive impact in various measures of student success such as retention and persistence rates, academic standing rates, graduation rates and student performance on measures such as grade point average. One respondent noted that there are too many variables to identify the specific impact of predictive modelling and interventions. A notable success of predictive models is that they have caused a shift in the culture and type of conversations taking place at some institutions and providing evidence for change, and the introduction of new supports and/or changes to how academic supports are promoted. As a result of predictive models, changes have been made to academic advising, student retention policies, curriculum design and educational scaffolding courses, orientation programs and first-year transition supports.

The most common challenges that respondents cite with the use of predictive modelling include issues with data (e.g., effort required to gather/clean data and data accuracy and timing), survey response rates, lack of resources (time, skills) and institutional acceptance. Advice that respondents would offer to institutions that had not yet embarked on predictive modelling emphasized the importance of engaging stakeholders and generating buy-in and being realistic about the extent of resources (e.g., knowledge, skills and time) required to build and maintain predictive models. The types of new data that institutions would like to incorporate into their models include academic performance, additional demographic information, student survey data, CRM information, use of various campus services or involvement in campus incidents, financial data and information from an LMS.

In many cases respondents felt that the current predictive model being used was meeting their needs. Where it was felt that the system was not meeting needs, the reasons cited included the fact that their predictive modelling was still in the early stages, lack of uptake of the modelling within their institution and the lack of time and/or resources. Most models are reviewed either regularly or as needed by in-house staff, sometimes with input from in-house faculty or system vendors.

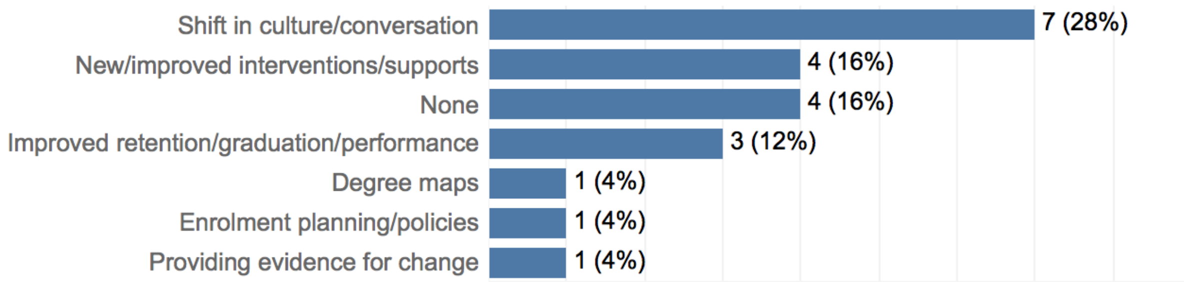
In an environment where evidence-based decision-making and allocation of resources is becoming more prevalent, it is clear that predictive modelling may play an important role in coming years as the models become more refined and mature over time.

Impacts of Using Predictive Modelling

Survey Questions 22 through 25 asked about the impact of predictive modelling on various measures of student success that institutions may track such as retention and persistence rates, academic standing rates, graduation rates and student performance on measures such as grade point average. In the majority of cases respondents either indicated that it was too early to see if there is an impact (between three and seven responses; 23–54%) or that they did not know/there was no impact (between two and six responses; 16–46%). Between one and five respondents (8 to 38%) indicated there was a positive impact. One respondent who indicated a positive impact noted increases in retention rates of approximately 5%, another noted a 2% overall increase with some groups, such as Aboriginal, seeing higher gains, while the remainder could not provide specific numbers. A respondent indicating a mix of positive and negative impacts noted that no systematic analysis had been conducted, but that anecdotally the models had flagged some students that would have been missed through other methods as well as some students that did not require interventions. Among the respondents who indicated a positive impact on graduation/completion rates, 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. The more detailed results for this set of questions can be found in Appendix C: Questions 22–25 and in Figures 32–35, and focus on retention/persistence, academic standing, graduation/completion rates, and performance such as grade point average, respectively.

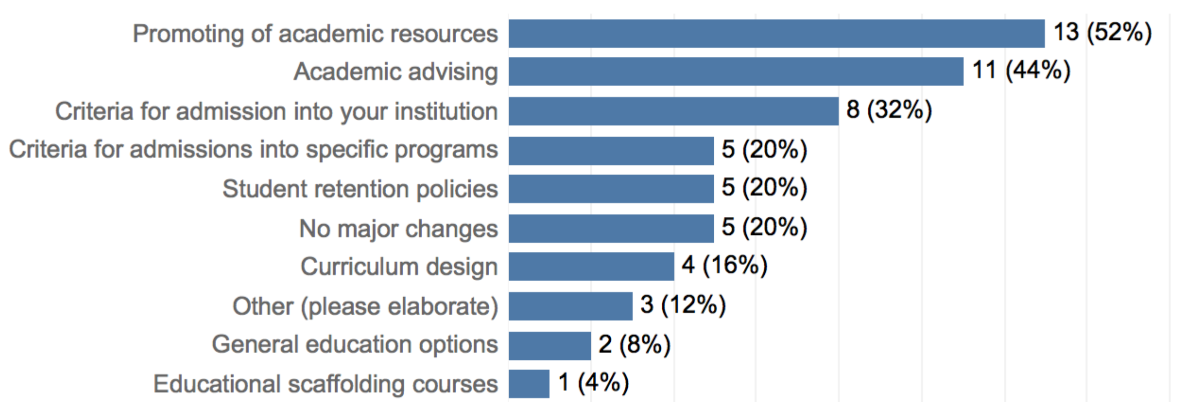
When asked about the biggest successes with the use of predictive modelling for student retention, the largest group of respondents (seven, 28%) indicated that use of predictive modelling had caused a shift in the culture and the conversation around the institution; respondents talked about "focusing the conversation on student retention" and how predictive modelling "got campus talking about student success"; related to this was one respondent (4%) who wrote about predictive modelling providing evidence for change at their institution, particularly with regard to interventions. Four respondents (16%) indicated that the creation or improvement of student interventions and supports was a big success, while three respondents (12%) saw improvements in student graduation, retention and performance. Degree maps — where the path a student must take through a program's curriculum is mapped out — and using evidence in enrolment planning and policies were mentioned by one respondent (4%) each. In addition to four respondents (16%) that indicated some version of "no successes," there were eight respondents (32%) that did not respond to this question (see Figure 19).

Figure 19: Responses to Question 27, "What are the biggest successes you've had with modelling student retention?" (Qualitatively Coded)



Question 32 asked what institutional changes had occurred because of the use of predictive modelling. As shown in Figure 20, 13 (52%) respondents using predictive modelling indicated changes to the way that academic resources were promoted, while changes to academic advising occurred at the institutions of 11 respondents (44%). Small numbers of respondents identified changes to admissions criteria at the institutional (eight, 32%) or program (five, 20%) level, changes to student retention policies (five, 20%), curriculum design (four, 16%), general education options (two, 8%), and educational scaffolding courses (one, 4%). Among the "other" group (three, 12%), two respondents identified changes to orientation and how students are transitioned into the institution and the other identified changes in the way that pass/fail decisions are made on particular students, but provided no additional information.

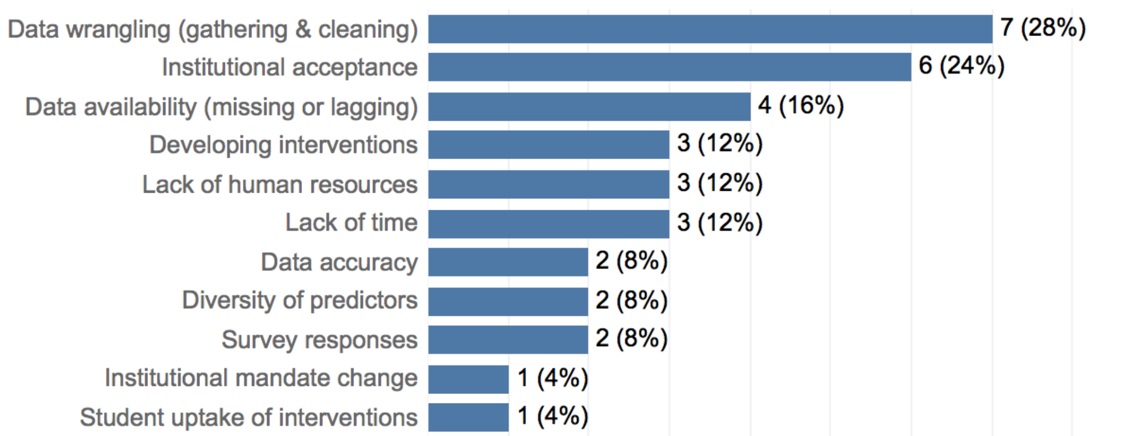
Figure 20: Responses to Question 32, "Has the use of predictive modelling at your institution lead to any major changes in the following areas?" (Select All That Apply)



Challenges

When asked about the challenges they had encountered in using predictive models for student retention, many of the survey responses talked about issues with data, such as the effort required to "wrangle" (gathering and cleaning) data (seven, 28%), dealing with missing data or data that takes a long time to become available (four, 16%), dealing with the diversity of elements that could be useful as predictors (two, 8%), the accuracy of data that is accessible (two, 8%), and getting students to complete surveys (two, 8%). Institutional acceptance (six, 24%) was another often-cited challenge, with respondents indicating difficulties with getting predictive models accepted and adopted within the community. A lack of resources was cited as a challenge for both time (three, 12%) and skilled people (three, 12%); developing interventions (three, 12%) and getting students to take up the interventions (one, 4%) were also mentioned. Six respondents did not answer this question.

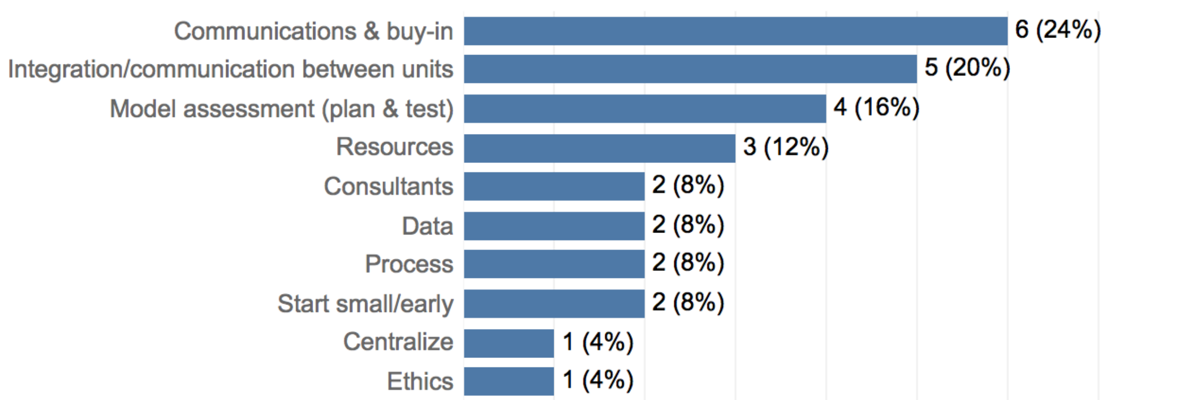
Figure 21: Qualitative Coding of Responses to Question 26, "What are the biggest challenges you've faced related to modelling student retention?"



The interview and questionnaire responses regarding challenges largely mirrored the responses to a similar question on the surveys — issues with data collection and cleaning, resourcing appropriately and knowledge levels were noted again. Two new items appeared as well, however, the first being the difficulty of dealing with large data sets and the computational requirements of predictive models, and how this can tax the existing information technology infrastructure. The other new item was the issue of "intuitiveness" of modelling, and how understanding the predictor variables selected by the model is not always straightforward or easy for model developers to explain to those that need to use the model.

When asked about the advice they would offer to institutions that had not yet embarked on predictive modelling for student retention, the largest group of respondents (six, 24%) suggested communicating with stakeholders and generating buy-in, with comments such as "articulate the reason why clearly," "it takes patience," and "it is important to build campus knowledge." In a related vein, five respondents (20%) talked about integration and communication between units, particularly from where the data is generated to where the modelling is done, and finally to the group(s) doing the interventions and other programming. Having a plan to regularly assess the predictive modelling was mentioned by four respondents (16%) and resourcing properly by three respondents (12%). Others suggested starting early (two, 8%), investigating and getting to know the data (two, 8%), having a process for building, testing and talking with stakeholders about the model (two, 8%), paying attention to and talking about the ethical dilemmas inherent in predictive modelling (one, 4%), and centralizing the building of the model (one, 4%). The use of external consultants came up in two responses, however in diametrically opposite ways: One respondent indicated that "external consultants can provide great support for developing the process for creating and socializing the model" while another simply said "don't trust consultants."

Figure 22: Qualitative Coding of Responses to Question 28, "What advice would you offer to an institution looking to implement predictive modelling for student retention?"



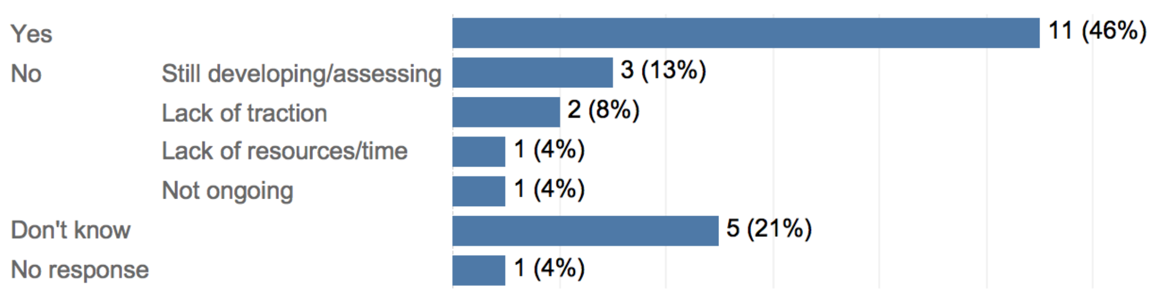
When asked what they wished they had known when they began predictive modelling, the nearly unanimous response from participants involved the knowledge and resourcing required to maintain and improve the model. In some cases, the concerns were technical resourcing, such as the right kinds of software or ensuring that the hardware available is adequate for the task, while in other cases it was personnel resourcing and understanding not only the time required but the required skills and knowledge. Two participants specifically mentioned beginning with outsourced models and not fully understanding the resource requirements when bringing those models under institutional control.

More than half of the respondents (13, 52%) indicated that there are types of data that are not currently used in their predictive modelling that they would like to incorporate. The types of data that could be used include academic performance, either at the secondary or postsecondary level (four), additional demographic information (four), student survey data (three), other institutional data, such as CRM information, or involvement in campus incidents (three), and financial data (two).

Assessing Whether Predictive Modelling Systems are Meeting Institutional Needs

Several survey respondents (11, 46%) felt that their current predictive modelling system and methods worked well for the needs of their institution (see Figure 23). A further five respondents (21%) did not know if their system worked well for their needs, while seven respondents (29%) indicated it did not; additionally, one respondent provided no response. Among those who did not feel their system and methods worked well, three respondents (13%) identified the fact that their predictive modelling was still in the early stages as their rationale, with two citing a lack of uptake of the modelling within their institution, one indicated a lack of resources and time was holding them back, and one responded that their predictive modelling was not yet an ongoing process.

Figure 23: Responses to Question 18, "Do you find that your current predictive modelling system and methods work well for your needs?" (Select One)

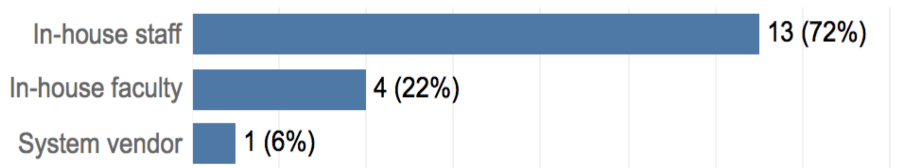


When asked if their system underwent regular reviews, responses were evenly split between indicating that their modelling had not been reviewed, that it was being reviewed annually, and that it was being reviewed regularly but not annually (see Figure 24).

Figure 24: Responses to Question 30, "Has your predictive modelling system been reviewed since it was originally implemented to see if refinements can be made?" (Select One)

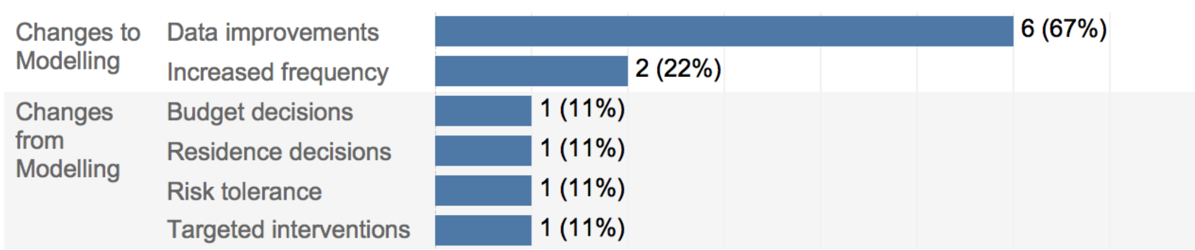
Where respondents indicated that their model had been reviewed regularly or annually, the subsequent question asked them who participated in the model review process. As shown in Figure 25, in-house staff was the most selected option with 13 responses (72% of 18 respondents asked this question), and a small number indicated utilizing in-house faculty (four, 22%) or the system vendor (one, 6%).

Figure 25: Responses to Question 31, "Is this model review an in-house process, or do you work with an external organization?" (Select All That Apply)



The question about how the use of predictive modelling has evolved since the initial implementation was interpreted in the interviews and questionnaires through two different lenses: Some responses spoke to how the modelling itself had changed, while others spoke to how the use of predictive modelling had changed the institutional context. Among the changes to modelling, responses included improvements to the data being used such as an increasing volume of data, number of predictors, or accuracy (six, 67%) and increases in the frequency with which the modelling is done (two, 22%). Changes from the modelling cited included an increase in risk tolerance based on the ability to model changes before implementation (one, 11%), and use of predictive modelling in residence decisions (one, 11%) and in targeting interventions (one, 11%). Additionally, one participant (11%) responded that the use of predictive modelling had provided more information and evidence to the institution's SEM community and that they helped pave the way for a budgeting model that was tied more closely to enrolments. (See Figure 26.)

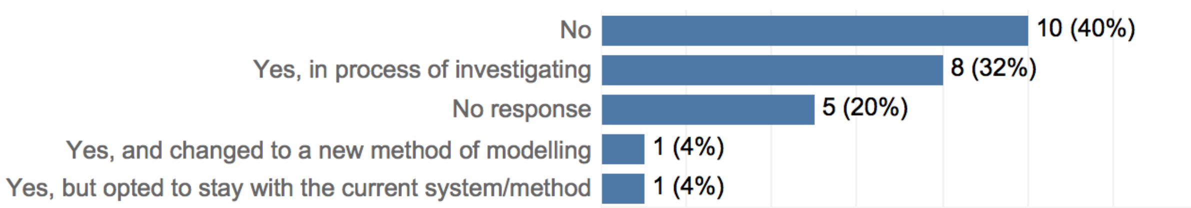
Figure 26: Phase Two Participants' Indication of Evolution of Predictive Modelling



Participants were also asked how they saw their institution's use of predictive modelling evolving over the next five years. The responses reinforced key themes identified in the survey — bringing new data into the modelling process and ensuring that the models are continually assessed and improved. Participants also indicated that they wished to see their predictive models used to make institutions more student centric, with one response specifically speaking to utilizing institutional money to incentivize changes toward that goal.

The final major question of the survey asked respondents about whether they have investigated alternative methods of predictive modelling compared to their current methods. As seen in Figure 27, the largest groups of respondents indicated that either they had not investigated alternatives (10, 40%) or they were currently in the process of investigating them (eight, 32%), while one respondent (4%) had investigated alternatives but chose to stay with their current method. One respondent (4%) had investigated and ultimately switched, and when asked to expand on that, indicated they had moved from a simplistic model to one that additionally took local unemployment rates into account. A further five respondents (20%) did not provide an answer to this question.

Figure 27: Responses to Question 33, "Have you investigated alternative methods of modelling student retention, such as using a different system or different predictive methods?" (Select One)



Discussion

This research into the use of predictive modelling at postsecondary institutions has revealed several overall themes and shed light on the research questions. In the discussion below, we start with looking at the current usage of predictive modelling at postsecondary institutions, and what practices and principles apply when using it. We then look at how predictive modelling is being used in innovative ways and leveraged to improve decisions, along with opportunities and challenges related to predictive modelling identified by respondents. Finally, we discuss future research directions and the limitations of this research.

Use of Predictive Modelling in Postsecondary

Of the 70 responses to our survey, 36% of respondents indicated they were using predictive modelling in some way within their institution and a further 39% responded that they were exploring or seriously considering implementation in the near future. Predictive models were also used in more than one context in many cases, such as using an aggregate model for enrolment or budget planning purposes, and a more individual model for interventions with at-risk students or promoting use of advising resources. The 68% of survey respondents who indicated they use predictive modelling for enrolment planning purposes also indicated they use predictive modelling for more student-specific purposes as well, and 44% of interview/questionnaire participants utilized both types of modelling.

When given the opportunity to indicate why they were using predictive modelling, many respondents indicated some sort of pressure on the institution or academic area. This pressure could be fiscal, as institutional budgets become tighter or local demographics shift, or it could be pressure to improve student success. While a number of respondents indicated that they were not pursuing predictive modelling because of data or resourcing issues, two survey respondents indicated the lack of uptake of predictive modelling was because it was not perceived as necessary, as retention and graduation rates at their institution were already high.

Adoption of predictive modelling has also been increasing in the last several years. Half of respondents that provided an answer to this question on the survey (11 of 25) indicated that their predictive models had been implemented between 2014 and 2017, and a further five indicated they had a predictive model in development.

This data suggests that predictive modelling in Canadian postsecondary education is in its infancy. It is clear that institutions using predictive modelling are still learning about the benefits and challenges associated with the modelling process, tools and results. Consequently, institutions looking to adopt predictive modelling are less likely to find either best practices or tools that can match their own needs. Further, 56% of respondents noted that modelling is an iterative process that is reviewed regularly or annually, suggesting that models will continue to be refined in the years ahead. Research notes that there are minimal empirically-informed resources that can help guide institutions as they establish predictive modelling (Gašević et al., 2016). This study, coupled with further research and inter-institutional sharing, will contribute to this knowledge base, particularly in Canada.

Predictive Modelling Practices and Principles

There were relatively few questions where agreement was clearly found between most or all respondents, with two exceptions. The first is agreement on resourcing predictive modelling appropriately, both in terms of the skill sets of the people tasked with developing and using predictive models and the time and technical infrastructures to support them. Predictive modelling requires specialized software toolsets, significant computational ability and access to often diverse data sources. Research on predictive modelling resourcing has been limited to date, but Gašević et al. (2016) cite the importance of informed and committed leadership as a key attribute of successful implementation.

The second area of agreement was around the security of predictive models. Access to predicted individual outcomes was very tightly limited to those who needed to know for the purposes of running the models, assessing the models or coordinating or performing interventions. While most respondents indicated there were no specific policies in place around predictive modelling, interviewees and questionnaire respondents spoke about the importance of treating predictions confidentially and as sensitive personal information under protection of relevant privacy legislation.

Gašević et al. (2016) prompted an expert panel with the question: "*For learning analytics to make a continued impact on learning and teaching, it would need to …*" and performed clustering analysis on the responses. They identified six clusters: data platform — standards and governance (including both sharing models and a guarantee of security); data use — accessible, transparent, valid/reliable; compatibility with existing values/practices/systems; strategy — whole-of-organization view; actionable tools with an evidential base; and supporting student empowerment. This shifts modelling from "something 'done to' educators and students, [to] something done with them in partnership" (Gašević et al., 2016, p.20).

Innovative Uses of Predictive Models

The broad strokes of predictive modelling for student retention were fairly similar across most respondents. Similar uses relied on similar techniques and approaches. However, several responses deserve note as being unique in some way, including the use of predictive modelling for spaces in student residences and making the results of predictive models available directly to students.

While a number of respondents indicated they used residence data as an input in their model — that is, whether or not a student was living in an on-campus or otherwise institution-maintained residence — only one institution indicated that they were using a predictive model to help inform who would be offered a place in residence. If a student failed to complete their first semester, there would be little to no opportunity to fill that residence space until the next residence application and offer cycle, often many months away; this, in turn, led to not only a loss of revenue for the institution but also a lost ability to provide a space to someone who could be aided by having a residence space.

Two respondents indicated that rather than keeping the results of the predictive modelling in the hands of faculty and staff of the institution, they made the information directly available to students. They coupled the predictions with information that allowed students to determine their own path forward by choosing what available supports might be of interest to them. As neither of these institutions opted in to phase two of the study, it is not possible to provide additional detail from these institutions on how this process works for students.

Some respondents indicated that their use of predictive modelling was connected with their institutional strategic enrolment management plan. One respondent in particular took this a step further and included their SEM committee in the annual assessment and review of their predictive modelling, along with an explicit approval process. Among the information presented to the committee are the variables chosen for the predictive model along with their relative weightings. The discussion allows for members of the committee to discuss elements that perhaps should be looked at for inclusion in the model and to see how the predictors change year to year, along with looking at past performance.

Opportunities and Challenges for Interventions

Utilizing predictive modelling to inform interventions with students is not straightforward; only 57% of survey respondents that have a predictive model use it alongside an intervention. In some cases, this was simply a matter of timing — respondents indicated that the plan was to connect the two, but that had not yet occurred — but there was also indication in both the survey and interview/questionnaire responses that there was a struggle with getting those responsible for interventions to see the value in the modelling. The other major challenge cited relates to the personnel with the requisite skills and time to wrangle data, assess predictors and develop models.

When asked about the impacts of predictive modelling and interventions, many respondents reported that it was either too early to tell or that they had not seen an impact on any of the various measures that might be used, such as retention, good academic standing, graduation rates or grade point averages. One respondent noted that there are too many variables to identify the specific impact of predictive modelling and interventions. This could be an opportunity to use appropriate experimental design procedures, similar to those noted in the literature review for the second phase of Mohawk College's predictive modelling project where students were divided into experimental groups and a control group, in order to better isolate the effectiveness of predictive modelling (Finnie et al., 2017b).

However, of those that did report an impact it was almost universally a positive one; there were no wholly negative impacts found, and the one respondent indicating both a positive and negative impact cited the fact that the model occasionally flagged people for intervention who did not truly require it. No predictive model can be correct 100% of the time, however, and in general the reported results were positive.

Many respondents indicated a shift in the culture and conversation on their campus as a result of using predictive modelling. This shift led toward the increased use of data and evidence in decision-making, particularly with regard to the supports provided and how students are made aware of the available supports. Respondents suggested that having several "wins" or improvements in which predictive modelling played a role helped with both the uptake of model information for interventions as well as with continuing the shift to data-informed decision-making.

Leveraging the Strengths of Predictive Modelling

A majority of survey respondents (15 of 25, 60%) indicated that the use of predictive modelling had led to some changes at their institution. In some cases these changes were around the availability and/or promotion of supports to students, while in other cases changes to curriculum design, admission requirements or retention policies were enacted. Where there was a known impact on student retention and success, that impact was nearly universally a positive one.

Among the interventions crafted by institutions, responses showed that there was no single way that institutions had made changes to go along with predictive modelling. In some cases, only changes to the way that supports were promoted to students was reported, while at least one respondent indicated a wholesale overhaul of the advising model for students was in progress. What is clear from the responses is that there is no predictive model that will work in each and every institutional context. Models need to be built based on the information available to them (both in terms of what information is available and how that information is structured) as well as in terms of what is going to be predicted. Since these elements often vary within and between institutions, building a predictive model must be grounded in the institutional context and is a process that typically requires considerable resources and attention.

Future Research Directions

One element not captured in the research is the question of how the use of predictive modelling affects student access, particularly for underrepresented groups. This would require a deeper dive into modelling outcomes and intervention uptake than was possible in the current study. In cases where predictive modelling had caused a shift in new student intakes, a pre- and post- analysis would be worthwhile as well.

Amongst respondents, most of the direct student interventions involve promotion of optional rather than mandatory services. The exact nature and content of the promotions could be important, as has been shown in research in behavioural economics on “nudges” (Thaler & Sunstein, 2008) that is starting to be applied to postsecondary education as well (Ross, White, Wright & Knapp, 2013).

This study identified that the number of institutions using predictive modelling has doubled in the last four years, and that many more are seriously considering predictive modelling in the near future. A future followup longitudinal study on the impact of predictive modelling would add to our understanding of the long-term effects on institutions and students. Additionally, further research in this area could help institutions identify how to effectively use predictive modelling with the information that they currently collect, and how to augment this with additional data in the future. It is important to note that a focus on data alone is unlikely to be sufficient: To be successful at changing an institution, modelling needs to be presented and contextualized so that it can drive organizational development (Macfadyen & Dawson, 2012).

Conversely, a large portion of the respondents indicated that they were not using predictive modelling (45 of 70 respondents, 64%, of responses to Question 4). A future study that explores why institutions might not be adopting or interested in adopting predictive modelling would aid in understanding barriers to adoption within Canadian higher education.

Lastly, because the majority of the survey employed in this study focused on institutions currently using predictive modelling, only minimal information was received from the group that is planning to implement it soon, investigating or seriously considering investigating predictive modelling. A richer set of responses from this group in terms of why they are moving toward predictive modelling, what barriers they face and what opportunities they see would help provide a more holistic understanding of the process of adopting predictive modelling.

Identified future directions for research found during the literature review include the development of ethics of predictive modelling for student retention (Gašević et al., 2016), creating an overall framework for discussing predictive modelling (Chatti, Dyckhoff, Schroeder & Thüs, 2013), and further research into interventions and how they affect students (Jayaprakash et al., 2014).

Limitations

There are some limitations to this research that must be acknowledged. First, with our focus on Canadian institutions, this research helps fill a void in terms of understanding predictive modelling within the Canadian context but has not allowed us to compare what Canadian institutions are doing with the wider international community.

Finally, it is important to note that this report is limited to the perspectives of those who voluntarily participated in the study and therefore, is not representative of Canadian institutions (and their staff). Rather, this report is meant to be exploratory and descriptive in nature, in order to provide a better understanding of the landscape of predictive modelling in Canada.

Conclusions

Predictive modelling is gaining popularity as a way to improve both institutional planning and student retention. Using a two-phase approach, including a survey to gather a breadth of responses followed by an interview/email questionnaire to delve more deeply into the experiences of those who have used predictive modelling at their institutions, we investigated the use of predictive modelling for student retention at Canadian institutions.

Among our respondents, the number of institutions using predictive modelling for student retention purposes more than doubled from 2013 to 2017, with institutions reporting changes to the way that interventions for students were promoted, delivered and assessed. Of the respondents, 36% were using predictive modelling, and another 39% reported seriously considering it. There is substantial breadth in the ways that predictive modelling is implemented — from many different techniques and data sources, to models for different groups of students, to how the information is made available to staff, faculty, and in two cases, students themselves.

One of the more innovative uses of predictive modelling was to influence which applicants would receive an offer of a space in residence. Additionally, there were a number of respondents who had specifically connected their use of predictive modelling to their strategic enrolment plans and local community needs; in one particular case, the SEM committee was actively involved in the annual review of the model.
This study leaves opportunities for future research. We did not explore how predictive modelling affects access, nor the effects of promoting mandatory rather than optional student support services. Because many of the models discussed in this report are relatively new, a followup study to determine longer range effects would be beneficial.

References

AACRAO (n.d.). *AACRAO - American Association of Collegiate Registrars and Admissions Officers*. Retrieved from http://www.aacrao.org.

AIR (n.d.). *Association for Institutional Research*. Retrieved from http://www.airweb.org/pages/default.aspx.

AMOSSHE (n.d.). *AMOSSHE - Home*. Retrieved from https://www.amosshe.org.uk.

ANZSSA. (2015). *Home*. Retrieved from https://www.anzssa.com.

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue. In S. Buckingham Shum, D. Gašević, & R. Ferguson (Eds.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267–270). doi:10.1145/2330601.2330666

ARUCC. (n.d.). *Welcome - ARUCC - Association of Registrars of the Universities and Colleges of Canada*. Retrieved from https://www.arucc.ca/en/.

Astin, A. W. (1993). *What matters in college?: Four critical years revisited*. San Francisco: Jossey-Bass.

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining, 1*(1), 3–17.

Bean, J. P., & Metzner, B. S. (1985). A conceptual model of nontraditional undergraduate student attrition. *Review of Educational Research, 55*(4), 485–540.

CACUSS. (n.d.). *Canadian Association of University and College Student Services*. Retrieved from https://www.cacuss.ca/index.html.

Campbell, J. P. (2007, May). *Utilizing student data within the course management system to determine undergraduate student academic success: An exploratory study* (Doctoral dissertation). Retrieved from ProQuest. (3287222.)

Campbell, J. P., deBlois, P. B., & Oblinger, D. G. (2007, July 6). Academic analytics: A new tool for a new era. *EDUCAUSE Review, 42*(4), 40–57. Retrieved from http://er.educause.edu/articles/2007/7/academic-analytics-a-new-tool-for-a-new-era.

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2013). A reference model for learning analytics. *International Journal of Technology Enhanced Learning, 4*(5–6), 318–331. doi:10.1504/IJTEL.2012.051815

CIRPA. (n.d.). *Home | CIRPA*. Retrieved from https://cirpa-acpri.ca.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement. 20* (1): 37–46. doi:10.1177/001316446002000104

Dangeti, P. (2017). *Statistics for machine learning*. Birmingham, UK: Packt Publishing.

Daniel, B. (2015). Big Data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology, 46*(5), 904–920. doi:10.1111/bjet.12230

Davis, C. M., Hardin, J. M., Bohannon, T., & Oglesby, J. (2007). Data mining applications in higher education. In K. D. Lawrence, S. Kudyba, & R. K. Klimberg (Eds.), *Data Mining Methods and Application* (pp. 123–148).

Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems, 49*(4), 498–506. doi:10.1016/j.dss.2010.06.003

Dolence, M.G. (1993). *Strategic enrollment management: A primer for campus administrators*. Washington, DC: American Association of Collegiate Registrars and Admission Officers.

Ellis, R. K. (2009). Field guide to learning management systems. ASTD Learning Circuits, 2009.

Finlay, S. (2014). *Predictive analytics, data mining and big data: Myths, misconceptions and methods*. London: Springer.

Finnie, R., Childs, S., & Wismer, A (2011). *Under-represented groups in postsecondary education in Ontario: evidence from the Youth in Transition Survey*. Toronto, ON: Higher Education Quality Council of Ontario.

Finnie, R., Fricker, T., Bozkurt, E., Poirier, W., Pavlic, D. (2017). *Using predictive modelling to inform early alert and intrusive advising interventions and improve retention*. Toronto, ON: Higher Education Quality Council of Ontario. [Referenced in paper as: Finnie et al., 2017a.]

Finnie, R., Fricker, T., Bozkurt, E., Poirier, W., Pavlic, D., Pratt, M. (2017). *Academic advising: Measuring the effects of "proactive" interventions on student outcomes*. Toronto, ON: Higher Education Quality Council of Ontario. [Referenced in paper as Finnie et al., 2017b.]

Gašević, D., Buckingham Shum, S., Nelson, K., Alexander, S., Lockyer, L., Kennedy, G., et al. (2016). *Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement*. Sydney, NSW: Australian Office of Learning & Teaching. Retrieved from http://www.olt.gov.au/system/files/resources/SP13_3249_Dawson_Report_2016.pdf.

Hoffmann, E., & Chamie, M. (2002). Standard statistical classifications: basic principles. *Statistical Journal of the United Nations Economic Commission for Europe*, *19*(4), 223–241.

Horvli, T. (2004). *Data warehouse presentation*. Retrieved from: https://pdfs.semanticscholar.org/presentation/1747/aa686c37ae3e4171e32a96b3eb996e366906.pdf

Hossler, D., & Bean, J. (1990). *The strategic management of college enrollments*. San Francisco, CA: Jossey-Bass.

Jayaprakash, S. M., Moody, E. W., Lauría, E. J., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics, 1*(1), 6–47.

Jia, P., & Maloney, T. (2015). Using predictive modelling to identify students at risk of poor university outcomes. *Higher Education, 70*(1), 127–149. doi:10.1007/s10734-014-9829-7

Kahu, E. R. (2013). Framing student engagement in higher education. *Studies in Higher Education, 38*(5), 758–773.

Kuh, G. D., Cruce, T. M., Shoup, R., Kinzie, J., & Gonyea, R. M. (2008). Unmasking the effects of student engagement on first-year college grades and persistence. *The Journal of Higher Education, 79*(5), 540–563.

Kuzilek, J., Hlosta, M., Herrmannova, D., Zdrahal, Z., & Wolff, A. (2015). OU Analyse: Analysing at-risk students at The Open University. *Learning Analytics Review, LAK15*(1), 1–16.

Lam, Y. L. J. (1984). Predicting dropouts of university freshmen: A logit regression analysis. *Journal of Educational Administration, 22*(1), 74–82. doi:10.1108/eb009886

Macfadyen, L. & Dawson, S. (2012). Numbers alone are not enough: why e-learning analytics failed to inform an institutional strategic plan. *Educational Technology and Society,* 15(3): 149–163.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. Retrieved from McKinsey Global Institute website: https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation.

Minaei-Bidgoli, B., & Punch, W. F. (2003). Using genetic algorithms for data mining optimization in an educational web-based system. In E. Cantú-Paz, J. A. Foster, K. Deb, L. D. Davis, R. Roy, U.-M. O'Reilly, et al. (Eds.), *Genetic and Evolutionary Computation — GECCO 2003* (Lecture Notes in Computer Science, vol. 2724, pp. 2252–2263). doi:10.1007/3-540-45110-2_119

Morris, L. V., Wu, S.-S., & Finnegan, C. L. (2005). Predicting retention in online general education courses. *The American Journal of Distance Education, 19*(1), 23–36. doi:10.1207/s15389286ajde1901_3

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students*. (K. Feldman, Ed.) (Vol. 2). San Francisco: Jossey-Bass.

QuestionPro. (n.d.) *QuestionPro Online Surveys*. Retrieved from https://ca.questionpro.com.

Reinard, J. (2006). *Communication Research Statistics.* Thousand Oaks, CA: Sage Publications.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353–387. doi:10.1037/a0026838

Ross, R., White, S., Wright, J., & Knapp, L. (2013). *Using behavioral economics for postsecondary success*. Cambridge, MA: ideas42.

Scalise, A., Besterfield-Sacre, M., Shuman, L., & Wolfe, H. (2000). First term probation: Models for identifying high risk students. In *Proceedings of the 30th annual Frontiers in Education Conference: Building on a century of progress in engineering education* (pp. F1F/11-F1F/16). doi:10.1109/FIE.2000.897696

SocioCultural Research Consultants. (2017). *Dedoose* (Version 7.6.21) [Web application software]. Retrieved from http://www.dedoose.com.

State University (2018). *College student retention – defining student retention, a profile of successful institutions and students, theories of student departure*. Retrieved from: http://education.stateuniversity.com/pages/1863/College-Student-Retention.html.

Statistics Canada. (2017, November 29). Education in Canada: Key results from the 2016 Census. *The Daily*. Retrieved from http://www.statcan.gc.ca/daily-quotidien/171129/dq171129a-eng.htm?HPA=1.

Thaler, R. H., Sunstein, C. S. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research, 45*(1), 89–125.

Tinto, V. (1987). *Leaving college: Rethinking the causes and cures of student attrition*. Chicago: University of Chicago Press.

University of Maryland University College. (2015). *Predictive analytics for student success: Developing data-driven predictive models of student success*. Retrieved from https://www.umuc.edu/documents/upload/developing-data-driven-predictive-models-of-student-success-final.pdf.

van Barneveld, A., Arnold, K. E., & Campbell, J. P. (2012). Analytics in higher education: Establishing a common language. *EDUCAUSE Learning Initiative, 1*, 1–11. Retrieved from https://library.educause.edu/~/media/files/library/2012/1/eli3026-pdf.pdf.

Wiggers, R., & Arnold, C. (2011). *Defining, measuring and achieving "student success" in Ontario colleges and universities* (@ Issue Paper No. 10). Toronto, ON: Higher Education Quality Council of Ontario. Retrieved from http://www.heqco.ca/SiteCollectionDocuments/AtIssueStudent%20Success%20ENG.pdf.

Zhang, Y., Oussena, S., Clark, T., & Kim, H. (2010). Use data mining to improve student retention in higher education - A case study. In *International Conference on Enterprise Information Systems* (pp. 190–197).

Higher Education Quality Council of Ontario

An agency of the Government of Ontario