



Conseil ontarien
de la qualité de
l'enseignement
supérieur

Un organisme du gouvernement de l'Ontario

Possibilités et défis de la modélisation prédictive pour la persévérance scolaire

Patrick Lougheed, Andrew Drinkwater
et Lynne Jamieson
Plaid Consulting



Publié par

Conseil ontarien de la qualité de l'enseignement supérieur

1, rue Yonge, bureau 2402
Toronto (Ontario) Canada M5E 1E5

Téléphone : 416-212-3893
Télec. : 416-212-3899
Web : www.heqco.ca
Courriel : info@heqco.ca

Citez cette publication dans le format suivant :

Lougheed, P., Drinkwater A. et Jamieson L. (2018) *Possibilités et défis de la modélisation prédictive pour la persévérance scolaire*, Toronto : Conseil ontarien de la qualité de l'enseignement supérieur.



Les opinions exprimées dans le présent document de recherche sont celles des auteurs et ne représentent pas nécessairement les opinions ou les politiques officielles du Conseil ontarien de la qualité de l'enseignement supérieur ou d'autres organismes ou organisations qui auraient pu apporter un soutien, financier ou autre, à ce projet. Imprimeur de la Reine pour l'Ontario, 2018.

Remerciements

Cette recherche a été appuyée par des douzaines d'intervenants partout au Canada et dans le monde. Nous tenons à exprimer nos remerciements les plus sincères à tous ceux qui ont participé à ce processus.

Tout d'abord, aux répondants de plusieurs établissements qui ont pris le temps de partager leurs expériences par le biais d'un sondage en ligne, et à ceux qui ont accepté de donner suite par une entrevue ou un courriel. Votre point de vue nous a aidés à mieux comprendre le paysage de la modélisation prédictive de la persévérance scolaire dans l'enseignement supérieur.

Deuxièmement, aux groupes de l'industrie qui nous ont aidés à faire en sorte que le message concernant cette recherche soit partagé à l'échelle mondiale, ce qui nous a aidés à attirer des répondants ayant une connaissance approfondie des fondements techniques, de l'utilisation et des plans futurs pour leurs modèles. Il s'agit notamment de l'Association des registraires des universités et collèges du Canada; l'Association des services aux étudiants des universités et collèges du Canada (ASEUCC); l'Association canadienne de planification et de recherches institutionnelles (ACPRI); l'Australia and New Zealand Student Services Association (ANZSSA); AMOSSHE, l'Organisation des services aux étudiants du Royaume-Uni; l'Association for Institutional Research (AIR) et l'American Association of College Registrars and Admissions Officers (AACRAO). Nous apprécions également les efforts des nombreux utilisateurs des plateformes de médias sociaux qui ont utilisé nos messages afin de nous aider à atteindre un public plus large, ainsi que tous ceux qui nous ont recommandé d'autres répondants potentiels.

Troisièmement, au Conseil ontarien de la qualité de l'enseignement supérieur (COQES) pour avoir financé cette recherche et nous avoir soutenus dans notre démarche. Nous avons grandement apprécié le soutien de Kate Hagerman et Lauren Hudak, qui ont travaillé avec nous en profondeur en nous fournissant des commentaires et des suggestions d'amélioration, et de Cassandra Cao et Helen Tewolde qui ont participé aux premières étapes du projet.

Sommaire exécutif

Aperçu du projet

La population du Canada est de plus en plus instruite, comme en témoignent l'augmentation des inscriptions aux études postsecondaires et l'obtention de diplômes postsecondaires (Statistique Canada, 2017). Toutefois, l'accès, la rétention et la réussite des étudiants des groupes sous-représentés continuent de poser un défi. Il est nécessaire d'identifier les étudiants à risque de ne pas réussir et de mettre à l'échelle les interventions afin de leur fournir un soutien utile. L'utilisation de modèles statistiques prédictifs (Finnie, Fricker, Bozkurt, Poirier et Pavlic, 2017) est une façon d'examiner ce problème. La modélisation prédictive, en général, est l'application de techniques de modélisation statistique et informationnelle comme la classification, la régression et l'apprentissage machine pour faire des prédictions basées sur des observations déjà enregistrées (Finlay, 2014).

Dans l'enseignement supérieur, la modélisation prédictive peut servir à la gestion des inscriptions, à l'amélioration des indicateurs de réussite des étudiants (p. ex., obtention d'un diplôme, rétention, moyenne générale, etc.), à la collecte de fonds et à l'obtention de nombreux autres résultats. La modélisation prédictive fait généralement appel à des données provenant de systèmes de gestion de l'apprentissage, de systèmes d'information sur les étudiants et d'enquêtes auprès des étudiants. De nombreux projets de modélisation prédictive se sont concentrés sur un seul établissement, mais il existe des exemples plus récents qui examinent des modèles utilisant des données provenant de plusieurs établissements. Dans le contexte canadien, le Mohawk College a mis en œuvre un système de modélisation prédictive en 2012 (Finnie et coll., 2017a), comme l'ont fait plusieurs autres établissements sondés pour cette étude.

En tant qu'étude exploratoire, le but premier de ce projet était de fournir un aperçu du paysage de la modélisation prédictive au Canada, et non un résumé des stratégies techniques sur la façon d'utiliser les modèles prédictifs¹. Le présent rapport met l'accent sur l'utilisation de la modélisation prédictive, en indiquant si, où et comment la modélisation prédictive est utilisée efficacement pour améliorer la réussite et la persévérance scolaire².

Grâce à des activités de sensibilisation ciblées auprès des établissements d'enseignement postsecondaire, les ministères et les personnes les plus susceptibles d'utiliser la modélisation prédictive ont été désignés pour participer au sondage en ligne. Certaines personnes ont également choisi de participer à des entrevues de suivi, à des entrevues approfondies et semi-structurées ou à des questionnaires (au choix du répondant) où elles pouvaient fournir plus de détails sur leur utilisation, leur contexte et leurs défis ainsi que sur toute mesure découlant de leur utilisation de modèles prédictifs.

1 L'annexe A et l'annexe C donnent un aperçu technique de l'utilisation de la modélisation prédictive.

2 Dans le contexte de ce projet, on entend par « persévérance scolaire » toute mesure de l'effectif ou du rendement des étudiants au-delà du point de la première inscription, comme la persévérance d'année en année, l'obtention d'un diplôme, le rendement mesuré par les notes moyennes, le rendement mesuré par le taux de réussite scolaire ou le rendement dans chaque cours.

Une approche à deux volets, comprenant l'envoi de demandes par l'entremise des listes d'envoi des groupes de l'industrie de l'enseignement supérieur et des courriels ciblés aux principaux intervenants des établissements canadiens, a été utilisée pour choisir les établissements qui recevraient le sondage. Au total, 170 réponses au sondage en ligne ont été recueillies. De ce nombre, 100 ont été exclues parce qu'aucune information n'avait été fournie, ou parce que des réponses d'essai ou des duplicatas de réponses avaient déjà été reçus; les réponses de l'étranger ont également été exclues, car leur nombre restreint rendait une analyse significative impossible. Il nous restait donc 70 réponses pour une analyse plus approfondie – dont 66 étaient complètes, et quatre étaient partiellement complètes. Des 66 réponses complètes, 38 provenaient d'universités et 28 de collèges et d'établissements polytechniques.

Après le sondage, on a demandé aux répondants s'ils étaient prêts à participer à une entrevue de suivi par téléphone ou à répondre à un questionnaire par courriel. Sept entrevues ont eu lieu et deux questionnaires par courriel ont été remplis.

Les questions de recherche pour ce projet étaient :

- a. Quels pratiques ou principes sont utilisés dans la modélisation prédictive dans le contexte postsecondaire?
- b. Quelles sont les utilisations novatrices des modèles prédictifs qui influent sur la persévérance scolaire?
- c. Quels sont les possibilités et les défis associés aux interventions éclairées par des modèles prédictifs?
- d. Comment les établissements d'enseignement postsecondaire peuvent-ils tirer parti des avantages de la modélisation prédictive pour améliorer la persévérance scolaire? Les résultats de la modélisation prédictive ont-ils modifié les pratiques ou les politiques de l'établissement et y a-t-il des données indiquant si ces changements ont permis d'améliorer l'accès, la rétention ou la réussite des étudiants?

Principales constatations

Les principales conclusions sont les suivantes :

- 36 % des répondants ont indiqué que leur établissement utilisait la modélisation prédictive pour la persévérance scolaire; 39 % ont indiqué que leur établissement étudiait, envisageait sérieusement ou prévoyait l'utiliser; 10 % n'utilisaient pas la modélisation prédictive et ne l'avaient pas envisagée; et 16 % avaient examiné la modélisation prédictive dans le passé, mais ne l'utilisaient pas actuellement et ne prévoyaient pas l'utiliser.
- Parmi ceux qui n'utilisaient pas la modélisation prédictive, les répondants ont indiqué que le ressourcement était un problème, que ce soit en matière de personnel, de temps ou d'outils, tandis que d'autres répondants ont mentionné des problèmes de qualité ou de compréhension des données.

- 52 % des répondants ont utilisé la modélisation prédictive pour éclairer des interventions spécifiques de persévérance scolaire. Parmi celles-ci, les interventions les plus courantes consistaient à promouvoir les services de soutien et à offrir de l'aide scolaire individuelle optionnelle.
- Certains répondants ont utilisé la modélisation prédictive dans le cadre de leur plan stratégique de gestion des inscriptions.
- Voici quelques-unes des utilisations les plus novatrices de la modélisation prédictive :
 - Un établissement a utilisé un modèle prédictif pour aider à déterminer quels demandeurs se verraient offrir une place en résidence.
 - Deux établissements ont déclaré qu'ils avaient mis à la disposition des étudiants les résultats des prédictions, ainsi que des renseignements qui leur ont permis de déterminer leur propre cheminement en choisissant parmi les mesures de soutien qui leur étaient offertes.
 - Un établissement a déclaré avoir fait participer son comité de gestion stratégique des inscriptions au processus de surveillance de son modèle prédictif. Cet examen permet au comité d'examiner les variables choisies pour le modèle, leur pondération relative et leur exactitude, et de discuter des points de données qui pourraient être pris en compte à l'avenir.
- Les répondants se sont entendus sur deux grands thèmes :
 - Les modèles prédictifs doivent être dotés de ressources adéquates, tant en matière de capital humain que d'infrastructure technique.
 - Les modèles prédictifs doivent être sécurisés de manière appropriée – en limitant l'accès à ceux qui ont besoin d'être informés de leur rôle dans la gestion ou l'évaluation des modèles, ou dans la coordination des interventions. Ces prédictions devraient être traitées comme tout autre renseignement personnel de nature délicate.
- Les défis de la modélisation comprennent la disponibilité des données, la transformation et la représentation cartographique de données, ainsi que les questions d'acceptation en établissements et de ressources.
- Parmi les réussites en matière de modélisation, mentionnons un changement de culture et de conversation, ainsi que des interventions et des soutiens nouveaux ou améliorés.
- Le conseil le plus souvent donné à ceux qui envisageaient de recourir à la modélisation prédictive était de mettre l'accent sur les communications et l'adhésion, ainsi que sur l'intégration et la communication entre les secteurs de l'établissement (p. ex. les facultés, les programmes universitaires, les départements administratifs).
- En partie à cause des pressions budgétaires et de la nécessité croissante de soutenir la réussite des étudiants, la majorité des systèmes de modélisation prédictive ont été mis en œuvre au cours des quatre dernières années et sont souvent établis à l'interne.

Conclusions et orientations futures de la recherche

Informar les activités d'intervention sur la base d'une modélisation prédictive est un défi – 52 % des répondants qui ont un modèle prédictif ont déclaré l'utiliser en parallèle aux interventions existantes. La trajectoire de la modélisation prédictive, en particulier dans l'enseignement supérieur canadien, n'en est encore qu'à ses débuts, plusieurs répondants affirmant que les interventions étaient sur leur feuille de route pour les deux prochaines années. D'autres ont laissé entendre qu'il était difficile d'amener ceux qui sont responsables des interventions à voir la valeur de la modélisation. L'une des plus grandes préoccupations soulevées par les répondants concernait le ressourcement, en matière de temps, de personnel et d'infrastructure.

Bien que de nombreux répondants aient estimé qu'il était trop tôt pour voir l'effet de la modélisation prédictive, ceux qui en avaient remarqué un l'ont qualifié de positif. Un effet positif particulier mentionné par de nombreux répondants est que la modélisation prédictive a aidé à faire évoluer la culture et la conversation sur leur campus vers une utilisation accrue des données et des éléments probants dans la prise de décisions, particulièrement en ce qui a trait à la prestation et à la promotion des services et du soutien aux étudiants. Le seul effet négatif noté par les répondants est que la modélisation a permis d'identifier certains étudiants qui autrement, auraient réussi (c'est-à-dire qui auraient réussi le cours, seraient restés en deuxième année, auraient obtenu un diplôme ou un autre résultat prévu). Comme les modèles ne permettent pas de voir l'avenir (Finlay, 2014), cet effet négatif peut être atténué en disposant de données plus nombreuses et de meilleure qualité pour améliorer la précision prédictive du modèle, mais il ne peut être complètement éliminé.

Il n'y avait pas de solution magique pour les établissements qui utilisaient la modélisation prédictive, car aucun des répondants n'a indiqué qu'ils utilisaient une solution « clé en main ». Les établissements ont plutôt fait remarquer que leur modèle dépendait des données dont ils disposaient et du contexte dans lequel ils fonctionnaient. Bien que certains établissements aient modifié leur façon de promouvoir les services aux étudiants, au moins l'un d'entre eux procédait à une révision complète de son modèle d'orientation universitaire fondé sur la modélisation prédictive.

Il ressort clairement de cette recherche que le contexte est un facteur important à la fois de l'orientation et du succès des efforts de modélisation. Dans certains cas, on a signalé que les mesures – les niveaux globaux d'inscription, les taux de rétention d'une année à l'autre ou les taux d'obtention du diplôme, par exemple – qui ont fait l'objet d'efforts de modélisation prévisionnelle initiale, étaient attribuables à des facteurs externes tels que les changements démographiques ou les changements de mandat des établissements. D'autres répondants ont indiqué que le principal facteur était la réussite des étudiants, qui est souvent influencée par des éléments comme la composition des étudiants desservis par l'établissement, les programmes et les cours offerts, et la structure des mesures de soutien qui leur sont offertes. Le succès des efforts de modélisation prédictive était le plus souvent lié à l'amélioration de la communication entre les différents secteurs de l'établissement et à l'amélioration de la capacité de celui-ci à prendre des décisions fondées sur des preuves.

Cette recherche a également permis de déterminer les orientations futures de la recherche, notamment l'incidence de l'utilisation de la modélisation prédictive sur l'accès des étudiants, l'incidence de la nature et du contenu de la promotion des interventions sur l'adoption de la modélisation prédictive et le suivi auprès du groupe qui met actuellement en œuvre la modélisation prédictive, tant à court terme pour ses plans que dans quelques années pour les leçons de mise en œuvre.

Table des matières

Possibilités et défis.....	1
de la modélisation prédictive pour la persévérance scolaire	1
Remerciements	2
Sommaire exécutif	3
Aperçu du projet.....	3
Principales constatations.....	4
Conclusions et orientations futures de la recherche.....	6
Table des matières.....	8
Liste des figures	10
Liste des tableaux.....	11
Introduction	12
Questions de recherche et méthodologie	14
Phase un : Sondage.....	15
Phase deux : Entrevue et questionnaire.....	16
Analyse.....	16
Revue de la littérature	17
Contexte général	17
Modélisation prédictive sans intervention.....	18
Interventions	19
Modélisation prédictive pour la persévérance scolaire au Canada	20
Résultats.....	22
Utilisation de la modélisation prédictive.....	25
Intervenants des établissements.....	30
Comment la modélisation prédictive est-elle utilisée?	32
<i>Types d'utilisation</i>	33
<i>Populations étudiantes</i>	34
<i>Données utilisées</i>	36
<i>Interventions</i>	40
Possibilités et défis de la modélisation prédictive.....	42

<i>Effets de l'utilisation de la modélisation prédictive</i>	43
<i>Défis à relever</i>	45
<i>Évaluer si les systèmes de modélisation prédictive répondent aux besoins des établissements</i>	47
Discussion.....	50
Utilisation de la modélisation prédictive dans les établissements postsecondaires	50
Pratiques et principes de la modélisation prédictive	51
Utilisations novatrices des modèles prédictifs	52
Possibilités et défis des interventions	53
Tirer parti des avantages de la modélisation prédictive	54
Orientations futures de la recherche	54
Limites.....	56
Conclusions	56
Références	57

Liste des figures

Figure 1 : Carte des réponses au sondage sur la modélisation prédictive pour la persévérance scolaire	24
Figure 2 : Carte des participants utilisant la modélisation prédictive à l'étape de l'entrevue et du questionnaire par courriel sur la persévérance scolaire	25
Figure 3 : Réponses à la question 4, « Votre établissement utilise-t-il actuellement une modélisation prédictive aux fins de la persévérance scolaire? » (Sélectionnez une réponse)	26
Figure 4 : Codage qualitatif des réponses autre que « Oui » à la question 4, « Votre établissement utilise-t-il actuellement une modélisation prédictive aux fins de la persévérance scolaire? »	27
Figure 5 : Réponses à la question 7, « Quand votre établissement a-t-il commencé à utiliser une modélisation prédictive aux fins de la persévérance scolaire? » (Sélectionnez une réponse)	28
Figure 6 : Réponses à la question 14, « Quelle a été l'impulsion de votre utilisation de la modélisation prédictive pour la persévérance scolaire? » (Sélectionnez toute réponse qui s'applique)	28
Figure 7 : Explication des participants à la phase deux pour l'adoption de la modélisation prédictive.....	29
Figure 8 : Indication des participants à la phase deux de l'origine de la demande de modélisation prédictive	30
Figure 9 : Réponses à la question 9, « Quand vous avez initialement mis en œuvre votre modèle prédictif pour la persévérance scolaire, qui a participé à sa mise en œuvre? » (Sélectionnez toute réponse qui s'applique).....	31
Figure 10 : Réponses à la question 16, « Qui a accès aux prédictions du système de modélisation prédictive? » (Sélectionnez toute réponse qui s'applique)	31
Figure 11 : Réponses à la question 17, « Comment les prédictions sont-elles mises à la disposition de ceux qui y ont accès? » (Sélectionnez toute réponse qui s'applique)	32
Figure 12 : Réponses à la question 5, « Vous avez indiqué que votre modèle prédictif est utilisé à des fins de persévérance scolaire. Comment votre établissement utilise-t-il la modélisation prédictive? » (Sélectionnez toute réponse qui s'applique).....	34
Figure 13 : Réponses à la question 13, « Pour quelles populations étudiantes votre établissement utilise-t-il des modèles prédictifs pour la persévérance scolaire? » (Sélectionnez toute réponse qui s'applique).....	35
Figure 14 : Réponses à la question 11, « De quels systèmes vos modèles prédictifs de persévérance scolaire recueillent-ils l'information? » (Sélectionnez toute réponse qui s'applique)	37
Figure 15 : Combinaison de réponses à la question 11, « De quels systèmes vos modèles prédictifs de persévérance scolaire recueillent-ils l'information? ».....	39
Figure 16 : Réponses à la question 12, « Quels types d'information vos modèles prédictifs utilisent-ils pour la persévérance scolaire? » (Sélectionnez toute réponse qui s'applique)	40
Figure 17 : Réponses à la question 20, « Utilisez-vous actuellement la modélisation prédictive pour orienter des interventions spécifiques en matière de persévérance scolaire? » (Sélectionnez une réponse).....	40

Figure 18 : Réponses à la question 21, « Veuillez indiquer les différents types d'interventions que votre établissement utilise en fonction de vos modèles prédictifs ». (Sélectionnez toute réponse qui s'applique) 41

Figure 19 : Réponses à la question 27, « Quelles sont les plus grandes réussites que vous avez eues en matière de modélisation de la persévérance scolaire? » (Qualitativement codé)..... 44

Figure 20 : Réponses à la question 32, « L'utilisation de la modélisation prédictive dans votre établissement a-t-elle entraîné des changements importants dans les domaines suivants? » (Sélectionnez toute réponse qui s'applique)..... 45

Figure 21 : Codage qualitatif des réponses à la question 26, « Quelles sont les plus grandes réussites que vous avez eues en matière de modélisation de la persévérance scolaire? » 46

Figure 22 : Codage qualitatif des réponses à la question 28, « Quels conseils donneriez-vous à un établissement qui cherche à mettre en œuvre un modèle prédictif pour la persévérance scolaire? » 47

Figure 23 : Réponses à la question 18, « Trouvez-vous que votre système et vos méthodes de modélisation prédictive actuels répondent à vos besoins? » (Sélectionnez une réponse)..... 48

Figure 24 : Réponses à la question 30, « Votre système de modélisation prédictive a-t-il été examiné depuis sa mise en œuvre initiale pour voir s'il est possible de l'améliorer? » (Sélectionnez une réponse)..... 48

Figure 25 : Réponses à la question 31, « Cet examen du modèle est-il un processus interne ou travaillez-vous avec une organisation externe? » (Sélectionnez toute réponse qui s'applique) 49

Figure 26 : Indication de l'évolution de la modélisation prédictive des participants à la phase deux..... 49

Figure 27 : Réponses à la question 33, « Avez-vous étudié d'autres méthodes de modélisation de la persévérance scolaire, comme l'utilisation d'un système différent ou de méthodes de prévision différentes? » (Sélectionnez une réponse)..... 50

Liste des tableaux

Tableau 1 : Taux de réponse au sondage par province/territoire 23

Tableau 2 : Réponses à la question 19, « Veuillez indiquer quel pourcentage de prédiction par rapport aux étudiants est exact dans votre modélisation : » 36

Introduction

La population du Canada est de plus en plus instruite, comme en témoignent l'augmentation des inscriptions aux études postsecondaires et l'obtention de diplômes postsecondaires (Statistique Canada, 2017). Toutefois, l'accès, la rétention et la réussite des étudiants des groupes sous-représentés continuent de poser un défi. Il est nécessaire d'identifier les étudiants à risque de ne pas réussir et de mettre à l'échelle les interventions afin de leur fournir un soutien utile. L'utilisation de modèles statistiques prédictifs (Finnie, Fricker, Bozkurt, Poirier et Pavlic, 2017) est une façon d'examiner ce problème.

L'objectif de cette recherche, effectuée au nom du Conseil ontarien de la qualité de l'enseignement supérieur (COQES) par Plaid Consulting, est de déterminer si, où et comment la modélisation prédictive est utilisée efficacement pour améliorer la persévérance scolaire afin d'être en mesure d'informer les futurs projets de modélisation prédictive et les nouvelles possibilités de tirer parti des données des établissements pour améliorer la réussite scolaire. Tel qu'indiqué ci-dessous dans l'analyse documentaire, peu de données ont été publiées sur la modélisation prédictive de la persévérance scolaire dans un contexte canadien, une lacune que cette recherche vise à combler.

En tant qu'étude exploratoire, le but premier de ce projet était de fournir un aperçu du paysage de la modélisation prédictive au Canada, et non un résumé des stratégies techniques sur la façon d'utiliser les modèles prédictifs³. Le présent rapport met l'accent sur l'utilisation de la modélisation prédictive, en indiquant si, où et comment la modélisation prédictive est utilisée efficacement pour améliorer la réussite et la persévérance scolaire⁴. Grâce à des activités de sensibilisation ciblées auprès des établissements d'enseignement postsecondaire, les ministères et les personnes les plus susceptibles d'utiliser la modélisation prédictive ont été désignés pour participer au sondage. Certains répondants ont également choisi de participer à des entrevues de suivi, à des entrevues approfondies et semi-structurées et à des questionnaires (au choix du répondant) où ils pouvaient fournir plus de détails sur leur utilisation, leur contexte et leurs défis ainsi que sur toute intervention découlant de leurs modèles prédictifs.

Le domaine de la modélisation prédictive n'est pas nouveau et son utilisation ne s'est pas limitée à un seul domaine ou à une zone géographique particulière. Des marchés financiers et du marketing aux équipes de baseball professionnel, la modélisation prédictive a trouvé de nombreuses applications différentes. L'enseignement supérieur n'est pas différent. Dans ce contexte, les modèles prédictifs sont des méthodes de plus en plus populaires qui tentent de prédire quels étudiants réussiront à certains égards, comme l'achèvement des cours, la rétention en deuxième année d'études, l'atteinte d'un certain seuil scolaire global ou l'achèvement du programme (Campbell, deBlois, & Oblinger, 2007). Une analyse plus détaillée de la modélisation prédictive se trouve à l'annexe A.

³ L'annexe A et l'annexe C donnent un aperçu technique de l'utilisation de la modélisation prédictive.

⁴ Dans le contexte de ce projet, on entend par « persévérance scolaire » toute mesure de l'effectif ou du rendement des étudiants au-delà du point de la première inscription, comme la persévérance d'année en année, l'obtention d'un diplôme, le rendement mesuré par les notes moyennes, le rendement mesuré par le taux de réussite scolaire ou le rendement dans chaque cours.

Le but de la modélisation prédictive est généralement de permettre des interventions ciblées afin d'influencer les résultats prévus. Il pourrait s'agir soit de se concentrer sur les résultats individuels positifs et d'intervenir pour s'assurer que ces résultats se produisent, soit de se concentrer sur les résultats négatifs et d'intervenir pour les transformer en résultats positifs, soit d'un mélange des deux options. Les modèles prédictifs ne peuvent en eux-mêmes ni prescrire des résultats ni des interventions – l'expertise humaine est nécessaire pour comprendre l'évaluation et choisir parmi les actions potentielles disponibles afin d'utiliser l'information du processus prédictif pour générer l'état futur souhaité.

Un point important concernant la modélisation prédictive, en particulier dans l'enseignement supérieur, est qu'il existe peu de solutions « clé en main » – celles qui peuvent être utilisées sans personnalisation avec les systèmes existants en place. Les solutions clé en main existantes utilisent généralement les données d'un système de gestion de l'apprentissage, ce⁵ qui peut être fait parce que le concept de réussite des cours est relativement standard dans l'enseignement supérieur. En dehors de l'achèvement du cours, les résultats escomptés, les données disponibles et le contexte (comme la conception du programme d'études, les caractéristiques démographiques de la population étudiante, les caractéristiques des établissements, etc.) varient généralement suffisamment d'un établissement à l'autre et même d'un programme à l'autre pour que toute modélisation prédictive doive être personnalisée. De plus, tous les systèmes doivent être revus et ajustés constamment pour tenir compte des changements dans les résultats, les données ou le contexte.

Cette personnalisation des modèles prédictifs dans l'enseignement supérieur signifie que pour savoir si, où et comment les modèles prédictifs sont utilisés et quels sont les défis et les opportunités qui en découlent, il faut obtenir des informations auprès de nombreux établissements différents. Il est possible qu'il existe de multiples modèles prédictifs au sein d'un établissement et nous devons recueillir de l'information auprès de plusieurs ministères et personnes au sein de ces établissements. De plus, à mesure que des modèles prédictifs sont utilisés à l'échelle mondiale, il est possible d'obtenir de l'information sur les pratiques exemplaires et les leçons apprises des établissements à l'extérieur du Canada.

La distinction entre ce qui constitue une analyse plus traditionnelle et le moment où cette analyse devient prédictive est au mieux grise et floue. Des techniques et des données sous-jacentes semblables, sinon identiques, sont utilisées, et les mesures des résultats sont souvent semblables. L'une des différences peut être le moment de l'analyse : Est-ce qu'un modèle est en train d'être créé pour expliquer des événements qui se sont déjà produits, ou est-ce qu'un événement futur est prévu?

⁵ Pour une description détaillée des systèmes de gestion de l'apprentissage, des exemples de systèmes et des données qu'ils contiennent habituellement, consultez le glossaire à l'annexe B.

Questions de recherche et méthodologie

L'objectif de cette recherche est de déterminer si, où et comment la modélisation prédictive est utilisée efficacement pour améliorer la persévérance scolaire afin d'être en mesure d'informer les futurs projets de modélisation prédictive et les nouvelles possibilités de tirer parti des données des établissements pour améliorer la réussite scolaire. Les questions de recherche spécifiques étaient les suivantes :

- a. Quels pratiques ou principes sont utilisés dans la modélisation prédictive dans le contexte postsecondaire?
- b. Quelles sont les utilisations novatrices des modèles prédictifs qui influent sur la persévérance scolaire?
- c. Quels sont les possibilités et les défis associés aux interventions éclairées par des modèles prédictifs?
- d. Comment les établissements d'enseignement postsecondaire peuvent-ils tirer parti des avantages de la modélisation prédictive pour améliorer la persévérance scolaire? Les résultats de la modélisation prédictive ont-ils modifié les pratiques ou les politiques de l'établissement et y a-t-il des données indiquant si ces changements ont permis d'améliorer l'accès, la rétention ou la réussite des étudiants?

Pour assurer une représentation exacte des utilisations des établissements de la modélisation prédictive dans la persévérance scolaire, il faut regarder au-delà de la documentation accessible au public. La modélisation prédictive en milieu collégial et universitaire est souvent entreprise par les membres du personnel (voir la figure 9, ci-dessous, pour les réponses à la question 9 qui demandait aux répondants qui participaient à la mise en œuvre des modèles prédictifs dans leur établissement) qui ne sont pas incités à publier de la même manière que les chargés de cours, ce qui a pour effet de rendre publique une information minimale sur la modélisation entreprise. La documentation disponible est souvent désuète, car les modèles sont constamment perfectionnés.

Afin d'obtenir un tableau plus complet, le projet a été structuré autour de multiples points de collecte de données. Premièrement, les ministères et les personnes les plus susceptibles d'utiliser la modélisation prédictive ont été invités à participer au sondage en communiquant avec les établissements postsecondaires du monde entier. Comme la modélisation prédictive est utilisée dans le monde entier, des réponses ont été recherchées auprès d'établissements du monde entier, l'accent étant mis en particulier sur les pays anglophones où la traduction ne serait pas nécessaire. Deuxièmement, certaines personnes ont également choisi de participer à des entrevues de suivi, à des entrevues approfondies et semi-structurées ou de répondre à des questionnaires (au choix du répondant) où elles pouvaient fournir plus de détails sur leur utilisation, leur contexte, leurs défis et les mesures découlant de leur utilisation de modèles prédictifs.

Phase un : Sondage

Afin de joindre le plus grand nombre de personnes possible avec le volet sondage de la recherche, nous avons fait appel à des groupes du milieu postsecondaire. Nous avons cerné des publics potentiels grâce à la recherche de groupes du milieu représentant les fonctions de recrutement, d'admission, d'orientation universitaire, de recherche institutionnelle et d'analyse dans les établissements postsecondaires, car ces unités d'affaires étaient les plus susceptibles d'utiliser la modélisation prédictive en ce qui a trait à la persévérance scolaire. Parmi les groupes du milieu postsecondaire identifiés, trois étaient axés sur le Canada, quatre sur les États-Unis avec quelques membres internationaux, trois couvraient d'autres pays anglophones (Australie, Irlande, Nouvelle-Zélande et Royaume-Uni) et un était mondial.

Une fois les groupes du milieu déterminés, nous avons communiqué avec les groupes potentiels pour leur fournir des renseignements généraux sur le sondage et des renseignements sur ce que nous leur demandions et dans quel délai. Nous avons d'abord identifié 11 groupes potentiels, dont huit ont accepté de participer, un a refusé et deux n'ont pas répondu. Une fois confirmés, les groupes de participants ont reçu des documents à envoyer à leurs membres, en leur demandant de les distribuer à la date de début du sondage. Un certain nombre de groupes canadiens ont utilisé leurs listes d'envoi de courriels aux membres pour annoncer le sondage⁶ et d'autres ont fait de la publicité au moyen de leurs bulletins d'information à l'intention des membres et de groupes de médias sociaux.⁷

En plus des efforts préliminaires de recrutement, 315 personnes provenant d'établissements canadiens qui n'avaient pas encore répondu à l'invitation à participer ont été contactées directement cinq semaines après le début du sondage. Cette mesure était complètement et délibérément distincte de la stratégie de recrutement initiale. Ces personnes peuvent ou non avoir reçu les invitations initiales par l'entremise des groupes du milieu, notamment les doyens ou vice-recteurs, les registraires, les directeurs des affaires étudiantes et les directeurs de la recherche institutionnelle. Les personnes contactées dépendaient des coordonnées disponibles sur le site Web de chaque établissement. Les personnes contactées qui n'ont pas répondu ont reçu un courriel de suivi après deux semaines, puis le sondage a pris fin deux semaines plus tard. Le sondage s'est tenu pendant un total de 9,5 semaines.

Le sondage a été effectué à l'aide de la plateforme de sondage QuestionPro hébergée au Canada (QuestionPro, s.d.). Le sondage initial a été élaboré en collaboration avec le COQES et demandait des renseignements aux représentants des établissements postsecondaires par rapport aux modèles prédictifs, notamment s'ils les utilisaient, quand et comment ils avaient été conçus, quelles interventions étaient fondées sur des modèles prédictifs, quelles étaient les répercussions des modèles et des interventions

6 Ces groupes comprenaient l'Association des registraires des universités et collèges du Canada, l'Association des services aux étudiants des universités et collèges du Canada (ASEUCC, s.d.) et l'Association canadienne de planification et de recherches institutionnelles (ACPRI, s.d.).

7 Ces groupes comprenaient l'Australia and New Zealand Student Services Association (ANZSSA, 2015); AMOSSHE, l'Organisation des services aux étudiants du Royaume-Uni; l'Association for Institutional Research pour les États-Unis avec quelques membres internationaux (AIR, n.d.); l'American Association of College Registrars and Admissions Officers, qui comprend également les États-Unis avec quelques membres internationaux (AACRAO, n.d.). Chaque groupe a fait une première annonce à la date de début du 8 août ainsi qu'un rappel par la même méthode deux semaines plus tard.

connexes, et quels défis et succès l'établissement avait constatés. L'instrument complet se trouve à l'annexe B. Pour minimiser le fardeau des répondants, seulement deux questions étaient obligatoires : le nom de l'établissement et si celui-ci utilise la modélisation prédictive pour la persévérance scolaire; pour les répondants qui ont indiqué utiliser la modélisation prédictive, nous leur avons demandé s'ils étaient prêts à participer à la deuxième phase de l'étude.

Au total, 170 réponses au sondage en ligne ont été recueillies. De ce nombre, 94 ont été exclues parce qu'aucune information n'a été fournie ou parce que des réponses ou des duplicatas de réponses avaient déjà été reçus. Six autres réponses ont été retirées de l'analyse, car elles provenaient d'établissements internationaux; bien que l'invitation à participer ait été lancée aux établissements internationaux, le peu de réponses obtenues a rendu impossible une analyse significative. Il ne restait donc que 70 réponses pour une analyse plus approfondie – dont 66 étaient complètes, et quatre étaient partiellement complètes. Le sondage a dépassé l'objectif initial de 40 réponses canadiennes.

Phase deux : Entrevue et questionnaire

Pendant la période de sondage, nous avons demandé aux répondants s'ils acceptaient de participer à la phase entrevue/questionnaire et lequel des deux ils préféraient. Tous les répondants au sondage ont eu le choix entre une entrevue ou un questionnaire. Les demandes d'entrevue ont été envoyées par courriel à 12 répondants, tandis que les demandes de questionnaire par courriel ont été envoyées à 10 répondants; les non-répondants de chaque groupe ont reçu un courriel de suivi. Des entrevues téléphoniques ont été menées auprès de sept personnes et des questionnaires ont été envoyés à trois répondants.

Les questions étaient les mêmes dans l'entrevue et dans le questionnaire, mais l'entrevue téléphonique a permis de clarifier certaines questions alors que ce n'était pas le cas pour les répondants au questionnaire. Les participants à cette phase ont été interrogés sur les mêmes thèmes que ceux du sondage, mais dans l'optique d'obtenir des détails et une profondeur supplémentaires qui n'étaient pas possibles avec le sondage. Par exemple, bien que le sondage ait été structuré de manière à fournir de l'information sur les grands types de données utilisés dans le processus de modélisation et sur l'impulsion donnée pour entreprendre la modélisation prédictive, l'entrevue et le questionnaire nous ont permis d'aller beaucoup plus loin en posant des questions de clarification. De plus, l'entrevue et le questionnaire nous ont permis de déterminer plus facilement des pratiques uniques en demandant des explications supplémentaires. L'instrument utilisé pour les entrevues et les questionnaires par courriel se trouve à l'annexe E.

Analyse

Les réponses aux questions ouvertes du sondage ainsi que les réponses au questionnaire par courriel et les notes d'entrevue ont été téléchargées dans Dedoose (SocioCultural Research Consultants, LLC, 2017) pour analyse qualitative. Avant le téléchargement, les données étaient rendues anonymes et le cryptage était activé dans Dedoose pour assurer les plus hauts niveaux de protection de la confidentialité disponibles. Le codage des réponses a été effectué par deux chercheurs qui ont codé respectivement 100 % et 38 % des

réponses aux questions individuelles; le *kappa* de Cohen (Cohen, 1960) était de 0,77. Les divergences notées dans le codage ont été corrigées par la discussion.

Revue de la littérature

Contexte général

La modélisation prédictive, en général, est l'application de techniques de modélisation statistique et informationnelle comme la classification, la régression et l'apprentissage machine pour faire des prédictions basées sur des observations déjà enregistrées (Finlay, 2014). S'appuyant sur les techniques d'exploration de données, la modélisation prédictive existe depuis des décennies sous diverses formes et est utilisée dans de nombreuses industries, des services financiers aux organismes sans but lucratif et au gouvernement (Finlay, 2014). Dans le domaine de l'éducation, la modélisation prédictive porte plusieurs noms, comme l'exploration de données éducatives (Baker et Yacef, 2009), l'analyse académique (Campbell et coll., 2007) et l'analyse d'apprentissage (Gašević et coll., 2016). Le terme utilisé dans la suite de ce rapport est celui de modélisation prédictive.

L'application de la modélisation prédictive dans l'enseignement supérieur dans le cadre de questions relatives aux résultats des étudiants, à savoir la réussite et la persévérance scolaire, peut en grande partie être attribuée à l'adoption de divers systèmes d'information éducatifs, tels que les systèmes de gestion de l'apprentissage (SGA), le système d'information sur les étudiants (SIE), les systèmes de gestion des relations avec la clientèle (CRM),⁸ et la tendance générale vers une utilisation accrue des médias sociaux et d'autres ressources technologiques. Ces systèmes fournissent aux établissements de grandes quantités de données qui peuvent être exploitées pour trouver des modèles et des prédicteurs (Daniel, 2015; Gašević et coll., 2016). La modélisation prédictive dans le domaine de l'enseignement supérieur peut être subdivisée en modélisation pour la gestion des inscriptions, la réussite des étudiants, la collecte de fonds ou bien d'autres résultats.

La réussite et la persévérance scolaire font depuis longtemps l'objet d'études dans le cadre de la recherche sur l'enseignement supérieur et orientent bon nombre des variables sous-jacentes utilisées dans les divers modèles prédictifs que l'on trouve dans la documentation. Ces modèles reposent généralement sur une combinaison de facteurs de rétention identifiés dans Tinto (1975, 1987), Bean et Metzner (1985), Astin (1993) ou Pascarella et Terenzini (2005). Ces facteurs de rétention peuvent comprendre des facteurs contextuels comme l'information démographique ou géographique, des facteurs scolaires comme les antécédents scolaires et le niveau de scolarité, l'intégration scolaire et sociale, et des facteurs externes comme les finances et les obligations professionnelles ou familiales. Des recherches plus récentes (Kuh, Cruce, Shoup, Kinzie et Gonyea, 2008) ont établi un lien entre la réussite et la participation des étudiants et

⁸ Pour une description détaillée et des exemples du SIE et du CRM, et des données qu'ils contiennent habituellement, veuillez consulter le glossaire à l'annexe B.

les activités éducatives utiles, comme les études et les activités sur le campus, et Kahu (2013) considère la rétention comme une conséquence distale de la participation. Finnie et coll. (2017a) ont identifié la « clarté de la carrière » et l'« engagement envers l'éducation » comme des indicateurs potentiellement significatifs de la probabilité qu'un étudiant abandonne ses études au Mohawk College.

Modélisation prédictive sans intervention

Les travaux de Lam (1984) sur la prévision des taux de décrochage des étudiants de première année à l'Université de Brandon à l'aide de techniques de régression logit et, dans la même veine, Scalise, Besterfield-Sacre, Shuman et Wolfe (2000) ont utilisé la régression logistique pour identifier les étudiants à haut risque en génie du premier trimestre à l'Université de Pittsburgh, constituent un exemple précoce de modélisation prédictive de la persévérance scolaire. Minaei-Bidgoli et Punch (2003), qui ont utilisé des algorithmes génétiques pour prédire les notes finales des cours pour les étudiants à partir des interactions des SGA, et l'analyse de Morris, Wu et Finnegan (2005) des données sur le rendement des écoles secondaires et des tests standardisés pour prédire la réussite des cours en ligne en utilisant une analyse discriminante prédictive, constituent d'autres exemples des premiers travaux. Les étudiants diplômés de l'Université de l'Alabama (UA) inscrits à un cours d'exploration de données en 2002 ont élaboré un modèle prédictif qui a permis d'identifier de 150 à 200 étudiants de première année chaque année qui ne reviendraient probablement pas pour leur deuxième année; ces données ont ensuite été communiquées aux chargés de cours et conseillers pour la sensibilisation et les interventions (Campbell et coll., 2007; Davis, Hardin, Bohannon & Oglesby, 2007). La combinaison de ces travaux prouve que les modèles prédictifs sont utiles pour la persévérance scolaire.

L'Université Purdue, J.P. Campbell (Campbell, 2007; Campbell et coll., 2007) a utilisé les données des SGA de Purdue pour créer un modèle de prévision de la réussite scolaire, tant pour la population générale que pour un modèle de première année seulement; ces modèles avaient des taux de réussite de 66 % et 80 % respectivement en étudiant des variables comme les notes au test d'habileté scolaire ou au ACT, la moyenne pondérée cumulative (MPC) et des variables composites représentant l'utilisation, l'évaluation et le calendrier des cours du SGA. Ce travail a par la suite été étendu au projet de « Course Signals » à Purdue.

Contrairement à d'autres projets, Jia et Maloney (2015) ont utilisé des données administratives seulement – c'est-à-dire qu'ils ont évité d'utiliser les données du SGA – d'une université néo-zélandaise pour prédire l'inachèvement de la première année et la non-retenue de la deuxième année en utilisant des modèles de risque prédictifs. Bien que les auteurs s'appuient davantage sur les modèles utilisés traditionnellement dans les domaines de la gestion des risques comme les soins de santé et la protection de l'enfance que sur les techniques d'exploration de données, il en résulte un modèle qui tient compte de nombreux facteurs semblables, comme la démographie et l'expérience scolaire antérieure. Leur modèle était beaucoup plus précis qu'un modèle non éclairé qui supposait que chaque inscription à un cours avait la même probabilité de terminer par un cours inachevé, avec les étudiants ayant obtenu le top 10 % des notes de risque les plus élevées représentant 29,55 % de cours inachevés en première année et 23,33 % d'abandons en deuxième année. Les auteurs n'ont étudié aucune intervention liée à leur modèle. L'une des principales leçons à

retenir est que différents types de modèles et d'analyses donnent des résultats tout aussi utiles lorsque la modélisation prédictive est appliquée à la persévérance scolaire.

Interventions

Arnold et Pistilli (2012) fournissent des renseignements sur l'évolution des Course signals après 2007. Les Course signals ont été mis en œuvre dans les SGA de Purdue, où les chargés de cours ont pu exécuter le modèle prédictif pour leurs étudiants en fournissant aux chargés de cours un indicateur de « feux de circulation ». Ce signal, à son tour, a été placé sur la page d'accueil du cours du SGA de l'étudiant, avec un signal vert pour ceux qui ont de fortes chances de réussite, jaune pour ceux qui peuvent avoir des problèmes et rouge pour ceux qui sont susceptibles d'avoir des difficultés. Le corps professoral décide ensuite d'interventions spécifiques auprès des étudiants, y compris des courriels, des messages texte, des renvois à des conseillers pédagogiques et des rencontres individuelles avec le chargé de cours. Les étudiants avec au moins un cours qui utilisait les Course signals avaient un taux de rétention significativement plus élevé que les étudiants qui n'utilisaient pas les Course signals, et les étudiants ayant deux cours ou plus avec les Course signals avaient un taux de rétention encore plus élevé. De plus, les taux de persévérance scolaire étaient plus élevés lorsque l'utilisation des Course signals s'effectuait plus tôt dans les études de l'étudiant. L'un des résultats de cette étude est la constatation que les interventions au niveau des cours peuvent avoir un impact positif sur la rétention au niveau des programmes et des universités.

L'Open University (OU) – un établissement d'enseignement à distance au Royaume-Uni – a également travaillé principalement sur les données du SGA et a mis au point un modèle prédictif pour identifier les étudiants à risque de ne pas réussir un cours, en commençant d'abord par deux cours d'introduction, puis 18 cours à partir de 2015 (Kuzilek, Hlosta, Herrmannova, Zdrahal, et Wolff, 2015). Le projet de l'OU a fait appel à plusieurs modèles de prévision différents et a utilisé à la fois des données démographiques et des données du SGA, puis est intervenu auprès des étudiants pour tenter de les ramener sur la bonne voie; les interventions prennent généralement la forme d'un appel téléphonique de la part des équipes spécialisées de soutien aux étudiants, mais Kuzilek et coll. ne fournissent aucune information sur l'efficacité des interventions. La principale constatation était l'importance cruciale de l'identification précoce des étudiants à risque de ne pas réussir un cours.

Alors que d'autres modèles prédictifs de la persévérance scolaire étaient axés sur un seul établissement, plusieurs projets ultérieurs ont porté sur la création de modèles utilisant des données provenant de plusieurs établissements. Dans un exemple, Jayaprakash, Moody, Lauría, Regan et Baron (2014) ont rendu compte de leurs travaux sur l'Open Academic Analytics Initiative (OAAI). L'OAAI a utilisé les données du SGA et les données de performance administrative du Marist College de l'État de New York pour construire un modèle prédictif de réussite des étudiants qui a ensuite été testé dans plusieurs collèges communautaires et, historiquement, dans des collèges et universités noirs. Le modèle a été utilisé pour informer les interventions auprès des étudiants dont on prédisait qu'ils risquaient d'avoir de la difficulté sur le plan scolaire. Les travaux de l'OAAI sont basés sur des logiciels ouverts et sont – du moins en théorie – transférables à d'autres contextes universitaires.

Les interventions étudiées par Jayaprakash et coll. (2014) comprennent des « messages de sensibilisation » basés sur les interventions par courriel utilisées par les Course signals à Purdue, et très semblables à ceux-ci, et la création d'un « online academic support environment (OASE) » (environnement de soutien scolaire en ligne) au sein du SGA de l'établissement qui fournit aux étudiants des ressources éducatives ouvertes. Les ressources du programme du OASE ont permis de faire connaître les mécanismes de soutien, les outils d'autoévaluation et les échafaudages pour améliorer les habitudes d'étude et rafraîchir le contenu fondamental. L'étude, menée dans tous les établissements participants à des cours sélectionnés (mais non identifiés) au printemps et à l'automne 2012 en utilisant un modèle créé et testé au Marist College à l'automne 2010 et au printemps 2011, a séparé les étudiants identifiés comme à risque sur le plan scolaire en un groupe témoin; un groupe « message de sensibilisation » et un groupe ayant accès au contenu du OASE. Les étudiants à risque sur le plan scolaire qui ont reçu des interventions ont généralement obtenu de meilleurs résultats que le groupe témoin (6 points de pourcentage de plus en moyenne), mais les groupes d'intervention ont généralement affiché des taux d'abandon de cours plus élevés que le groupe témoin. L'abandon anticipé peut en fait être un résultat positif puisque, si les étudiants croient ne pas être en mesure d'améliorer suffisamment leur note pour réussir, l'abandon peut leur permettre d'éviter un effet négatif sur leur relevé de notes. Cette étude suggère que les modèles et les interventions peuvent être utilisés dans tous les contextes tout en conservant une grande partie de leur pouvoir.

Une autre initiative à établissements multiples a été menée au University of Maryland University College (UMUC), qui a entrepris un projet de quatre ans visant à cibler les facteurs déterminants qui pourraient prédire la réussite des étudiants après leur transfert de deux collèges communautaires du Maryland (University of Maryland University College, 2015). Ils ont utilisé les observations des facteurs démographiques, des habitudes de fréquentation et du rendement des collèges communautaires, de la participation aux cours en ligne et du rendement en début d'études à l'UMUC pour prédire le rendement futur, la réinscription (inscription au deuxième trimestre), la rétention (inscription dans les 12 mois suivants le premier trimestre) et l'obtention du diplôme à l'UMUC. L'UMUC a constaté que différents facteurs étaient prédictifs de résultats différents – par exemple, le sexe était prédictif du rendement au premier trimestre, de la réinscription et de la rétention, mais non de l'obtention du diplôme, tandis que le fait qu'un étudiant ait suivi un cours de mathématiques au collège communautaire avant de passer à l'UMUC était prédictif du rendement au premier trimestre, mais non de l'obtention du diplôme, de la réinscription ou du maintien. L'UMUC a également étudié quatre interventions différentes effectuées sur les étudiants après leur transfert du collège communautaire à l'UMUC. Dans trois des interventions – l'utilisation d'une liste de contrôle des ressources pour les étudiants, le mentorat et un cours d'intégration préalable à l'inscription – n'ont pas montré de différences statistiquement significatives dans la moyenne pondérée cumulative, la réussite du cours ou la réinscription entre les groupes témoins et expérimentaux, tandis que la dernière intervention utilisant le tutorat en personne a révélé des différences significatives. Il convient de noter la petite taille des groupes expérimentaux, de sorte que ces résultats doivent être traités avec prudence.

Modélisation prédictive pour la persévérance scolaire au Canada

Dans le contexte canadien, Finnie et coll. (2017a) rendent compte des travaux menés au Mohawk College, où la recherche et la mise en œuvre d'un système de modélisation prédictive ont débuté en 2012. À l'aide

de données provenant d'étudiants qui sont allés au Mohawk College entre 2005 et 2012, le projet a créé des modèles prédictifs pour les étudiants qui sont entrés à l'automne 2013 et à l'automne 2014 et les a classés selon les catégories de risque des étudiants faible, moyen ou élevé, ce qui représentait leur probabilité de rétention au deuxième semestre et à la deuxième année d'études. Les variables du modèle comprenaient des renseignements sur l'étudiant et son programme, les réponses à un sondage lors de l'admission et les résultats aux évaluations en lecture, en écriture et en mathématiques effectuées avant le premier semestre de l'étudiant. Les résultats ont montré que 9 % des étudiants à faible risque qui sont entrés en 2013 et 2014 ont quitté Mohawk College, 13 % des étudiants à risque moyen et 23 % des étudiants à risque élevé ont abandonné; le modèle s'est révélé assez précis dans les prévisions. Bien qu'à ce stade, le modèle prédictif n'ait pas été jumelé à des interventions pour les étudiants, les auteurs notent qu'à mesure que le niveau de risque des étudiants augmentait (tel que déterminé par le modèle), les étudiants étaient plus susceptibles de chercher des ressources de consultation au moins une fois pendant leur première année au Mohawk College.

Dans la deuxième phase du projet du Mohawk College, les nouveaux étudiants de l'automne 2015 ont été affectés au hasard à l'un des trois groupes suivants : un groupe témoin qui a reçu un courriel avant le début des cours, un groupe expérimental qui a reçu jusqu'à trois autres courriels et un appel téléphonique encourageant la participation à une séance de consultation collective, et un groupe expérimental semblable qui a utilisé des conseils individualisés. Bien que l'affectation de groupe n'ait pas été orientée par le modèle prédictif de Finnie et coll. (2017a), Finnie et coll. (2017b) rendent compte des effets du traitement sur les groupes expérimentaux par les catégories de risque des étudiants du modèle prédictif. Les résultats ne sont pas particulièrement clairs, bien qu'ils montrent que pour certains groupes, il y a eu des améliorations statistiquement significatives des taux d'abandon. Les étudiants à risque élevé du groupe avec séances de consultation étaient 4,8 points de pourcentage moins susceptibles de partir ($p=0,1$) après leur premier trimestre, et les étudiants à risque faible du même groupe étaient 4,9 points de pourcentage moins susceptibles de partir ($p=0,1$) après leur première année, les hommes dans ce groupe étant 9,8 points de pourcentage moins susceptibles de partir ($p=0,05$).

En résumé, les travaux antérieurs sur la modélisation prédictive de la persévérance scolaire s'étendent au-delà des frontières, avec des travaux en Amérique du Nord, en Europe et en Océanie. Bien que l'on ait beaucoup mis l'accent sur les données du SGA comme principale source de données, d'autres modèles se concentrent soit sur l'utilisation de données administratives seulement, soit sur la collecte directe de renseignements auprès des étudiants. Peu des projets de recherche identifiés ont été transformés en systèmes opérationnels continus, les systèmes propriétaires ayant plutôt été déplacés dans l'espace; des exceptions notables sont au University of Alabama où le système décrit a été utilisé pendant au moins cinq ans, le projet des Course signals de Purdue et du système de l'Open University.

Dans la documentation disponible sur la modélisation prédictive de la persévérance scolaire, très peu se concentre sur le contexte canadien. Les travaux de Lam (1984) ont utilisé des données sur les étudiants de l'Université de Brandon, tandis que ceux de Finnie et coll. (2017a, 2017b) ont fait état de travaux au Mohawk College. Le manque de documentation faisant intervenir les établissements canadiens a été l'un des principaux facteurs à l'origine de ce projet.

Résultats

Le sondage comprenait une page d'accueil où les répondants potentiels pouvaient trouver de l'information sur les raisons pour lesquelles le projet était entrepris, un lien vers la politique de confidentialité qui régissait la collecte de l'information et les coordonnées de Plaid Consulting et du COQES si les répondants avaient des questions. Après avoir cliqué sur le bouton « Suivant », le logiciel du sondage enregistrait une réponse au sondage; au total, 170 réponses ont été enregistrées.

De ces 170 réponses, nous en avons exclu 100 de notre bassin de réponses à examiner et à analyser : dans 58 cas, on a cliqué sur le bouton « Suivant », mais aucune information n'a été fournie; dans six cas, les réponses provenaient de l'extérieur du Canada; dans cinq cas, le répondant a entré des données dénuées de sens, comme le nom d'un établissement « test » ou « ghjg »; dans 15 cas, le répondant a entré un nom d'établissement sans répondre à la question 4 demandant si l'établissement utilise la modélisation prédictive pour retenir ses étudiants; dans quatre cas, le répondant a répondu « oui » à la question 4, mais n'a pas répondu à la question 5 qui demandait comment il utilisait la modélisation prédictive; et dans 12 cas, la réponse était une répétition d'une réponse que nous avons incluse dans l'analyse. Il nous restait donc 70 réponses utilisables pour une analyse plus approfondie – dont 66 étaient complètes, et quatre étaient partiellement complètes. L'exhaustivité indique que les répondants ont appuyé sur la touche « Soumettre » à la dernière page du sondage, et comme le sondage était linéaire et sans méthode pour sauter des questions, nous savons qu'elles ont toutes été consultées, mais ne comprennent pas nécessairement des réponses. Les réponses partielles fournissent de l'information jusqu'aux questions 9, 18, 19 et 21 du sondage, respectivement. Les réponses partielles ont été incluses dans l'analyse, car un certain nombre d'autres réponses, bien que complètes, n'ont pas répondu à certaines des questions du sondage, et nous ne voyions aucune raison d'exclure le groupe partiellement complet simplement parce que leurs questions sont restées sans réponse à la fin du sondage.

Dans certains cas, des réponses multiples ont été reçues de différents secteurs d'un même établissement, au total, 70 réponses ont été reçues de 66 établissements canadiens; les 12 réponses en double retirées étaient les cas où elles provenaient clairement de la même personne dans le même secteur de l'établissement. Des réponses ont été reçues d'établissements de neuf des dix provinces (à l'exception de Terre-Neuve-et-Labrador) et le Yukon était le seul des trois territoires du pays à être représenté. L'Ontario a enregistré le plus grand nombre de réponses avec 26, suivie de la Colombie-Britannique (18) et de l'Alberta (11). Le tableau 1 ci-dessous indique le nombre d'établissements répondants par province ou territoire du Canada et un taux de réponse approximatif pour chacun. Il est à noter que ce tableau tient compte des établissements plutôt que des réponses individuelles, alors que le nombre total d'établissements dans

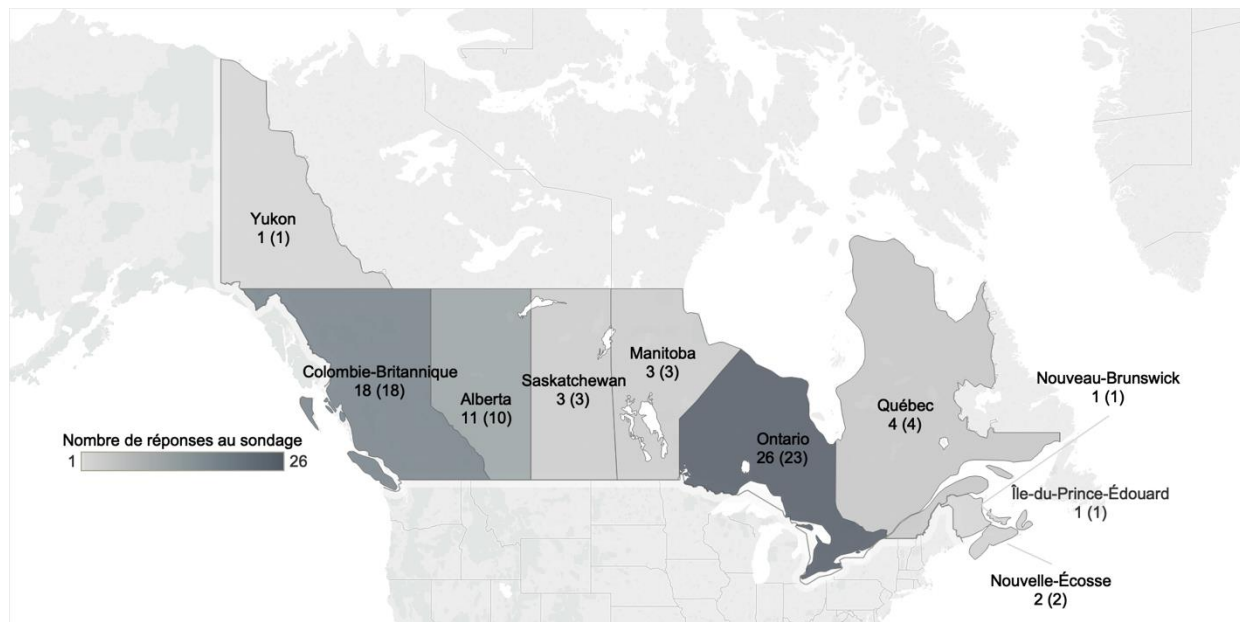
chaque province ou territoire comprend les établissements d'enseignement postsecondaire financés par les fonds publics et les établissements membres de Collèges et instituts Canada, de Polytechnics Canada ou d'Universités Canada.

Tableau 1 : Taux de réponse au sondage par province/territoire

Province/Territoire	Nombre d'établissements répondants	Nombre total d'établissements	Taux de réponse
Alberta	10	20	50 %
Colombie-Britannique	18	27	66,7 %
Manitoba	3	8	37,5 %
Nouveau-Brunswick	1	7	14,3 %
Terre-Neuve-et-Labrador	0	4	0 %
Territoires du Nord-Ouest	0	2	0 %
Nouvelle-Écosse	2	10	20 %
Nunavut	0	1	0 %
Ontario	23	54	42,6 %
Île-du-Prince-Édouard	1	3	33,3 %
Québec	4	68	5,9 %
Saskatchewan	3	15	20 %
Yukon	1	1	100 %
Total	66	220	29,1 %

Voir la figure 1 pour une carte des réponses par province ou territoire. Dans certains cas, des réponses multiples ont été reçues de différentes parties d'un même établissement, et le premier chiffre indiqué est le nombre de réponses reçues, tandis que le deuxième chiffre entre parenthèses est le nombre d'établissements distincts ayant répondu.

Figure 1 : Carte des réponses au sondage sur la modélisation prédictive pour la persévérance scolaire



Remarque : Une couleur plus foncée indique plus de réponses; les chiffres représentent le nombre de réponses au sondage, et ceux entre parenthèses indiquent le nombre d'établissements distincts ayant répondu.

La plupart des réponses au sondage proviennent de personnes ayant une responsabilité de surveillance ou un rôle dans l'élaboration du modèle prédictif. La figure 28 de l'annexe C présente la ventilation détaillée. À la fin du sondage, on a demandé aux 25 répondants des établissements utilisant la modélisation prédictive s'ils étaient prêts à participer à une entrevue de suivi par téléphone ou à répondre à un questionnaire par courriel. Sept entrevues ont eu lieu et deux questionnaires par courriel ont été remplis par les participants de la phase deux de la collecte de données.

La figure 2 illustre les participants de cette phase; contrairement à la phase un, chaque établissement n'a fourni qu'une seule réponse, et la coloration et la numérotation représentent à la fois le nombre de réponses et le nombre d'établissements. Comme les questions posées aux participants au cours des entrevues téléphoniques étaient essentiellement semblables à celles fournies aux participants dans les questionnaires par courriel, l'analyse a porté sur les réponses aux deux dans leur ensemble. Les questions initiales pour les deux méthodes étaient les mêmes, bien que l'entrevue ait permis de poser des questions supplémentaires afin de clarifier ou d'approfondir certains points.

Figure 2 : Carte des participants utilisant la modélisation prédictive à l'étape de l'entrevue et du questionnaire par courriel sur la persévérance scolaire

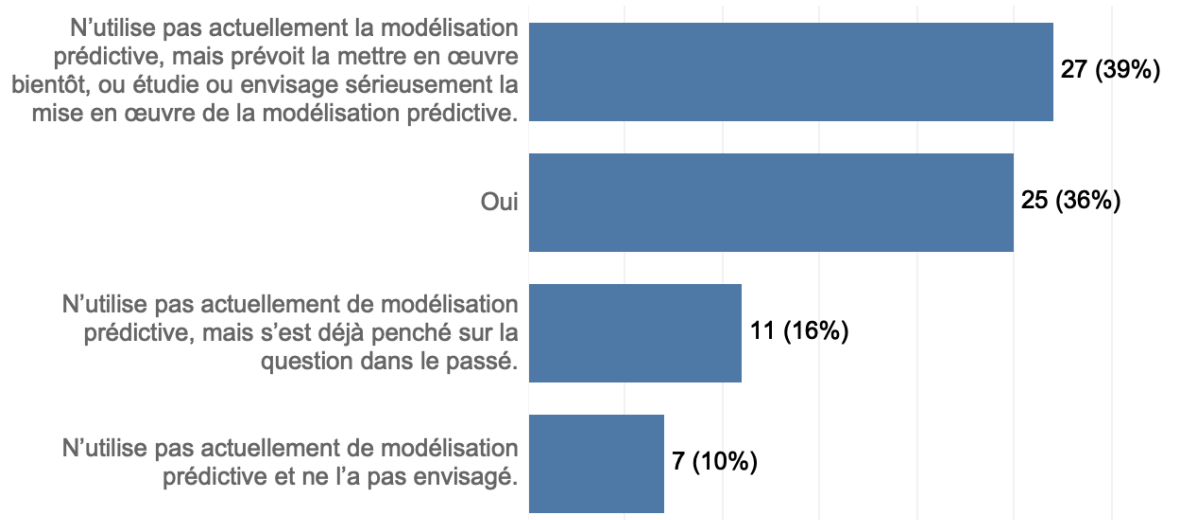


Utilisation de la modélisation prédictive

Les résultats du sondage ont démontré que l'utilisation de la modélisation prédictive dans les établissements d'enseignement postsecondaire à des fins de persévérance scolaire est en hausse, particulièrement depuis 2013. La plupart des établissements qui utilisent ou envisagent d'utiliser la modélisation prédictive font remarquer que l'impulsion de le faire était d'appuyer la réussite des étudiants, un plus petit nombre d'entre eux citant les exigences ou les priorités des établissements. L'information recueillie lors des entrevues et/ou dans les questionnaires a montré que dans presque tous les cas, la modélisation prédictive est utilisée à l'échelle de l'établissement plutôt qu'au sein d'un département ou d'un programme particulier, et que les utilisations les plus courantes sont la planification des inscriptions et les projections des admissions. Parmi les autres raisons invoquées pour justifier la mise en œuvre de la modélisation prédictive, mentionnons l'amélioration de la réussite des étudiants et l'identification des étudiants à risque de ne pas réussir sur le plan scolaire. L'utilisation de la modélisation prédictive est généralement encouragée par les cadres supérieurs de l'établissement ou des bureaux responsables de la planification et des services d'inscription.

Le groupe final d'analyse des réponses comprenait 70 réponses au sondage. De ce nombre, 25 (36 %) ont indiqué que leur établissement utilise actuellement la modélisation prédictive aux fins de la persévérance scolaire et 27 (39 %) ont mentionné qu'ils l'étudient, l'envisagent sérieusement ou prévoient l'utiliser. Ces résultats sont présentés plus en détail à la figure 3 ci-dessous.

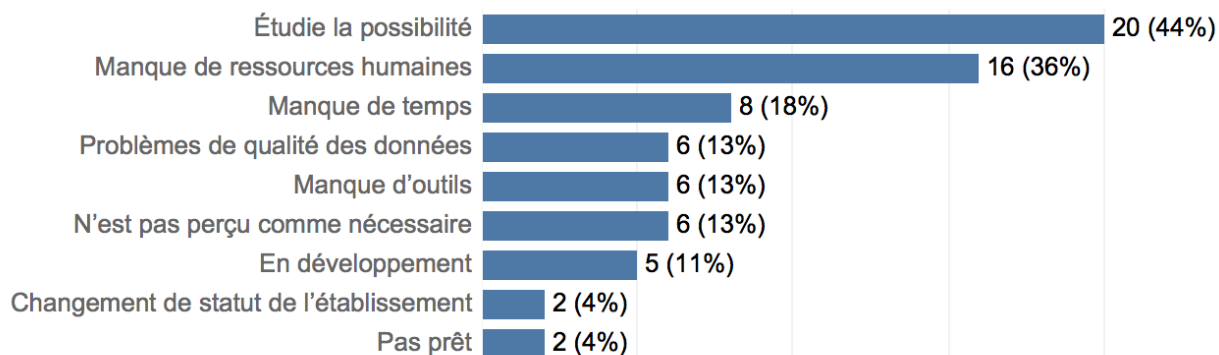
Figure 3 : Réponses à la question 4, « Votre établissement utilise-t-il actuellement une modélisation prédictive aux fins de la persévérance scolaire? » (Sélectionnez une réponse)



Les répondants qui ont répondu « oui » à la question 4 ont eu l'occasion d'expliquer pourquoi ils n'utilisaient pas actuellement la modélisation prédictive et s'ils pensaient que cela pourrait changer dans l'avenir. Ces réponses ont été codées qualitativement et les résultats sont présentés à la figure 4. De ces 45 répondants, 20 (43 %) ont indiqué qu'ils exploraient actuellement la possibilité d'utiliser la modélisation prédictive, mais qu'ils n'étaient pas encore sur le point de mettre définitivement en œuvre un système. Un certain nombre de réponses ont indiqué que le ressourcement était un problème, soit en matière de personnel (16, 35 %), de temps (huit, 17 %) ou d'outils comme des logiciels appropriés et du matériel suffisamment puissant (six, 13 %), tandis que six répondants (13 %) ont indiqué que la qualité des données ou la compréhension les retenaient.

Une autre préoccupation, citée à deux reprises (4 %), était un changement dans le mandat de l'établissement, car les changements apportés aux programmes qui en découlent signifient que la rétention historique peut ne pas être prédictive de la rétention future, et que ces établissements peuvent devoir attendre plusieurs années avant que la modélisation prédictive puisse être raisonnablement poursuivie. Six répondants (13 %) ont indiqué que la modélisation prédictive n'était pas perçue comme un besoin dans leur établissement, et trois d'entre eux ont cité des taux de rétention et d'obtention de diplôme déjà élevés comme étant la raison pour laquelle ils ne l'avaient pas fait; deux autres (4 %) ont indiqué que l'établissement n'était pas prêt pour la modélisation prédictive, sans indiquer pourquoi. Enfin, cinq répondants (11 %) ont indiqué que la modélisation prédictive irait de l'avant et qu'ils en étaient actuellement à l'étape de l'élaboration, l'un d'eux indiquant un projet pilote pour l'automne 2017 et un autre pour l'automne 2018.

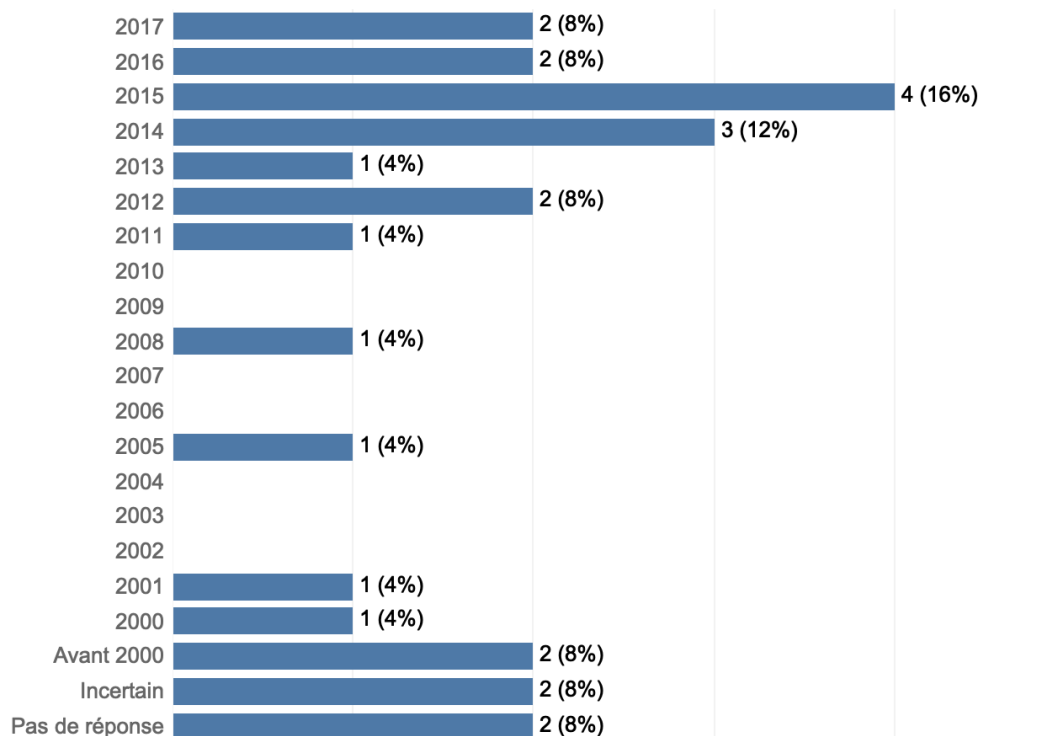
Figure 4 : Codage qualitatif des réponses autre que « Oui » à la question 4, « Votre établissement utilise-t-il actuellement une modélisation prédictive aux fins de la persévérance scolaire? »



Les 25 répondants qui ont indiqué à la question 4 que leur établissement utilisait la modélisation prédictive ont été dirigés vers le reste du sondage, tandis que les autres répondants ont été dirigés directement à une page les remerciant de leur participation au sondage. L'analyse qui suit porte sur les 25 répondants qui disaient utiliser la modélisation prédictive.

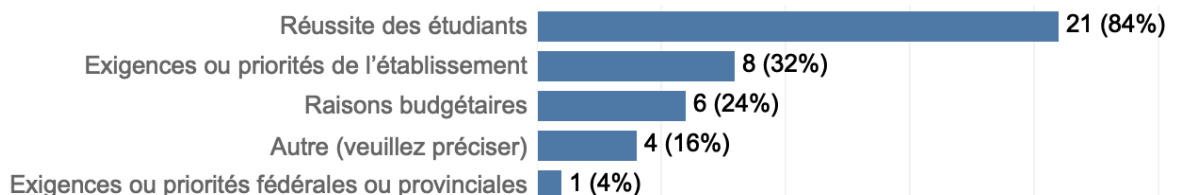
L'utilisation de la modélisation prédictive est un phénomène relativement récent dans les établissements d'enseignement postsecondaire, et bon nombre des réponses montrent que la modélisation prédictive n'est en place que depuis 2013 ou plus tard, la période 2014-2017 ayant suscité 11 réponses (39 %). Avant 2013, l'adoption de la modélisation prédictive s'est effectuée à un rythme beaucoup plus lent (voir la figure 5).

Figure 5 : Réponses à la question 7, « Quand votre établissement a-t-il commencé à utiliser une modélisation prédictive aux fins de la persévérance scolaire? » (Sélectionnez une réponse)



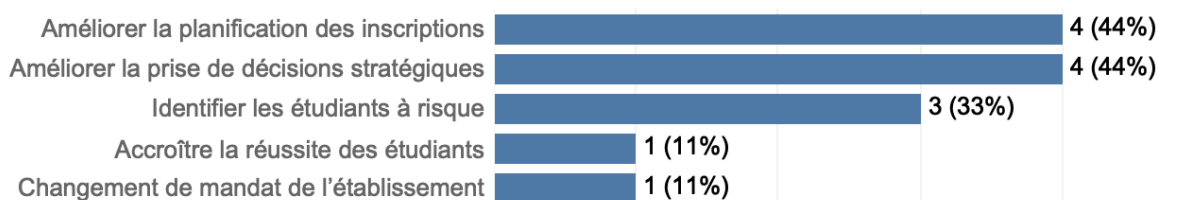
Lorsqu'on leur a demandé pourquoi ils ont commencé à utiliser la modélisation prédictive, la plupart des établissements (21, 84 %) ont indiqué qu'ils avaient commencé pour des raisons de réussite des étudiants, un plus petit nombre citant les exigences ou les priorités des établissements (8, 32 %), des raisons budgétaires (six, 24 %) ou des exigences ou priorités fédérales ou provinciales (un, 4 %). Parmi ceux qui ont répondu « autre », l'un a indiqué la planification des inscriptions, l'autre l'amélioration des résultats scolaires, l'un l'adaptation des interventions pour la réussite des étudiants et l'autre l'amélioration du programme et la compréhension des besoins futurs des étudiants. (Voir la figure 6.)

Figure 6 : Réponses à la question 14, « Quelle a été l'impulsion de votre utilisation de la modélisation prédictive pour la persévérance scolaire? » (Sélectionnez toute réponse qui s'applique)



Ces questions ont été examinées plus en détail avec les neuf établissements qui ont participé à une entrevue ou rempli un questionnaire (voir la figure 7). Les raisons les plus souvent citées pour lesquelles les participants ont commencé à utiliser la modélisation prédictive comprenaient l'amélioration de la planification des inscriptions (quatre, 40 %), l'amélioration de la prise de décisions stratégiques (quatre, 40 %) et l'identification des étudiants à risque de ne pas réussir sur le plan scolaire (trois, 30 %). Un participant (10 %) a indiqué qu'il s'agissait d'améliorer la réussite des étudiants. Un établissement (10 %) a indiqué que le changement de mandat et les changements connexes apportés aux programmes rendaient invalide la méthode longitudinale historique (selon laquelle de nombreuses années de données historiques ont été utilisées pour évaluer les années futures) et que de nouvelles techniques étaient nécessaires pour interpréter les données sur des périodes beaucoup plus courtes.

Figure 7 : Explication des participants à la phase deux pour l'adoption de la modélisation prédictive

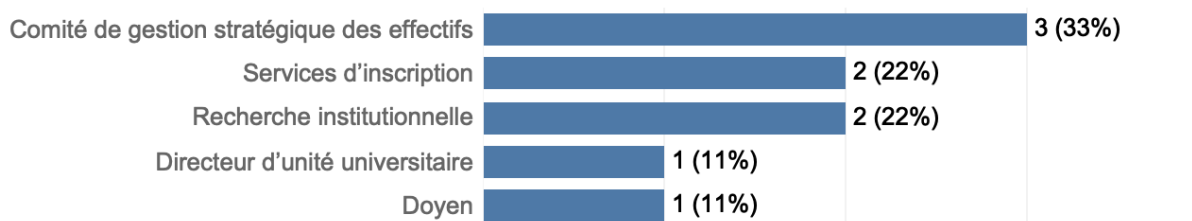


Remarque : Certaines réponses ont reçu plusieurs codes.

Presque tous les participants au sondage et aux entrevues (huit sur neuf, 89 %) ont déclaré utiliser la modélisation prédictive dans l'ensemble de l'établissement plutôt qu'au sein d'un département ou d'un programme particulier.

La modélisation prédictive a été effectuée sur une base ad hoc et régulière en fonction de l'objectif du modèle et de la disponibilité des données. Les participants ont déclaré avoir des modèles qui étaient à la fois agrégés – c'est-à-dire qui prévoyaient la rétention pour un groupe d'étudiants sans prévoir la rétention pour un étudiant en particulier – et individuels. Ces résultats sont présentés plus en détail à l'annexe C (voir la figure 37).

La demande de commencer à utiliser la modélisation prédictive provenait le plus souvent des cadres supérieurs de l'établissement, comme un comité de gestion stratégique des inscriptions (trois, 33 %) ou le bureau du recteur (un, 11 %), des bureaux responsables de la planification, comme le bureau de recherche de l'établissement (deux, 22 %), ou des services des inscriptions (deux, 22 %). Dans un cas (11 %), la demande provenait du directeur de l'unité d'enseignement et c'était le cas lorsque la modélisation n'a été faite que pour un programme particulier plutôt que pour l'ensemble de l'établissement (voir la figure 8).

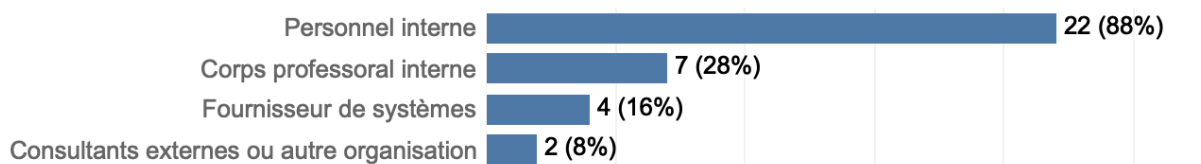
Figure 8 : Indication des participants à la phase deux de l'origine de la demande de modélisation prédictive

Intervenants des établissements

La plupart des systèmes de modélisation prédictive actuellement utilisés dans les établissements d'enseignement postsecondaire ont été mis en œuvre par le personnel interne ou le corps professoral. Dans certains cas, on a fait appel à un fournisseur de systèmes, à un consultant externe ou à une autre organisation. Les résultats de la modélisation prédictive ont été mis à la disposition principalement des gestionnaires administratifs, des conseillers universitaires, d'autres membres du personnel administratif, des directeurs d'unité, des comités de gestion stratégique des inscriptions et d'autres administrateurs supérieurs. Peu de répondants ont indiqué que les membres du corps professoral et les chargés de cours non membres du corps professoral y avaient accès. Dans deux cas, les résultats ont été mis directement à la disposition de l'étudiant. Dans la plupart des cas, les prévisions réelles du système ont été rendues disponibles par le biais de rapports personnalisés.

Parmi les établissements qui utilisent actuellement la modélisation prédictive, 22 des 25 répondants au sondage (88 %) ont indiqué que leurs systèmes étaient mis en œuvre par du personnel interne. Dans certains cas, des membres du corps professoral à l'interne y ont également participé (sept répondants, soit 28 %). Dans quatre cas (16 %), on a eu recours au fournisseur de systèmes et, dans deux cas (8 %), à un consultant externe ou à une autre organisation (figure 9). Parmi ceux qui ont choisi plus d'une des options de réponse, six (24 %) ont indiqué que le système avait été mis en place par le personnel interne, les chargés de cours internes et le fournisseur du système. Trois autres (12 %) ont indiqué que le système avait été mis en place par le personnel interne et le fournisseur du système. Aucune autre combinaison n'a obtenu plus d'une réponse, et un répondant n'a fourni aucune réponse à cette question. Lors de la sélection des « consultants externes », les répondants ont eu la possibilité de donner le nom de l'organisation du consultant, deux d'entre eux indiquant Noel-Levitz.

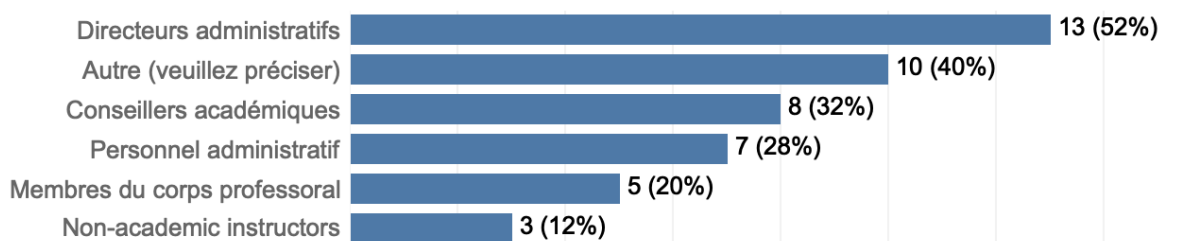
Figure 9 : Réponses à la question 9, « Quand vous avez initialement mis en œuvre votre modèle prédictif pour la persévérance scolaire, qui a participé à sa mise en œuvre? » (Sélectionnez toute réponse qui s'applique)



Lorsqu'on a demandé qui a accès aux résultats de la modélisation prédictive, le sondage a montré que la majorité des systèmes donnent accès à certains gestionnaires administratifs (13, 52 %), tandis que dans huit cas (32 %), l'accès est fourni aux conseillers universitaires et dans sept autres (28 %) à d'autres membres du personnel administratif. Peu de systèmes donnent accès aux membres du corps professoral (cinq, 20 %) ou aux chargés de cours non membres du corps professoral (trois, 12 %).

Les 10 (40 %) réponses dans la catégorie « autres » comprenaient : les directeurs d'unité d'enseignement (deux réponses), le comité de gestion stratégique des inscriptions (deux), les cadres supérieurs (deux), l'étudiant (deux) directement et au cas par cas (trois). Ces résultats sont présentés plus en détail à la figure 10.

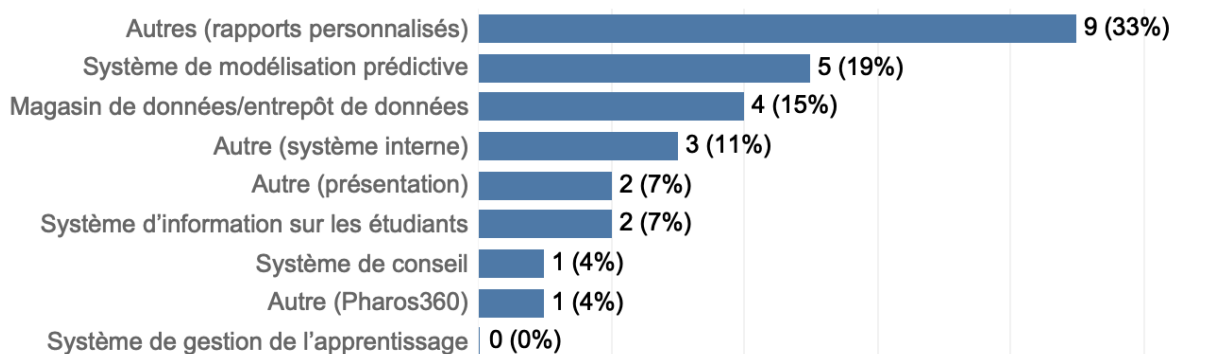
Figure 10 : Réponses à la question 16, « Qui a accès aux prédictions du système de modélisation prédictive? » (Sélectionnez toute réponse qui s'applique)



Lorsqu'on leur a demandé comment ils gardaient la sécurité des prédictions en s'assurant que seules les personnes autorisées pouvaient avoir accès à l'information, les participants aux entrevues et aux questionnaires ont répondu qu'en général, l'accès à l'information lorsque les prédictions étaient faites sous forme globale était moins restreint que si les prédictions étaient faites à l'échelle individuelle. Dans deux cas (20 %), les prédictions ont été téléchargées sur un système opérationnel – dans un cas sur un système CRM et dans l'autre sur le SIE – où la sécurité appropriée a été appliquée. Dans d'autres cas avec des données individuelles, les informations n'étaient stockées que dans une base de données à accès extrêmement limité. Lorsque les prédictions sont globales, l'information a plus souvent été mise à la disposition des groupes de l'ensemble de l'établissement, mais il fallait s'assurer que l'information n'était pas transmise à l'extérieur de l'établissement.

Les résultats du sondage ont montré que, dans la plupart des cas, les prévisions réelles du système sont disponibles soit par l'intermédiaire du système lui-même (cinq, 29 %), soit par le biais de rapports personnalisés (neuf, 33 %) comme le montre la figure 11. Un plus petit nombre de réponses a été observé pour le magasin de données ou l'entrepôt de données (quatre, 15 %), le système d'information sur les étudiants (deux, 7 %) et le système d'aide (un, 4 %); l'option du SGA n'a pas été choisie. Dans le groupe « autre », trois répondants (11 %) ont indiqué un autre système, deux répondants (7 %) ont indiqué que les prédictions sont fournies dans les présentations et, dans un cas (4 %), par le système Pharos360 mentionné précédemment.

Figure 11 : Réponses à la question 17, « Comment les prédictions sont-elles mises à la disposition de ceux qui y ont accès? » (Sélectionnez toute réponse qui s'applique)



Remarque : Total des réponses = 30; réponses « Autres » séparées par un codage qualitatif.

Comment la modélisation prédictive est-elle utilisée?

Les modèles prédictifs sont utilisés pour de nombreuses raisons associées à l'amélioration de la réussite des étudiants, notamment l'identification des étudiants vulnérables, le ciblage des interventions, la promotion du soutien aux étudiants et l'amélioration de la planification des inscriptions. Les étudiants de première année de premier cycle constituent le principal groupe cible des modèles. Les perceptions des répondants varient en ce qui concerne l'exactitude des prédictions pour différents groupes d'étudiants. Les données utilisées dans les modèles prédictifs sont le plus souvent collectées à partir du SIE de l'établissement. Les données peuvent également être utilisées à partir d'autres systèmes comme des systèmes pour l'aide financière, les conseils, l'engagement des étudiants et la de gestion de l'apprentissage. Les types d'information couramment utilisés dans les modèles prédictifs comprennent les données démographiques sur les étudiants, l'emplacement, les antécédents scolaires, les sondages auprès des étudiants, les tests normalisés et les renseignements sur les admissions et les demandes. Plus de la moitié des répondants ont indiqué qu'ils utilisent des modèles prédictifs pour éclairer des interventions spécifiques en matière de

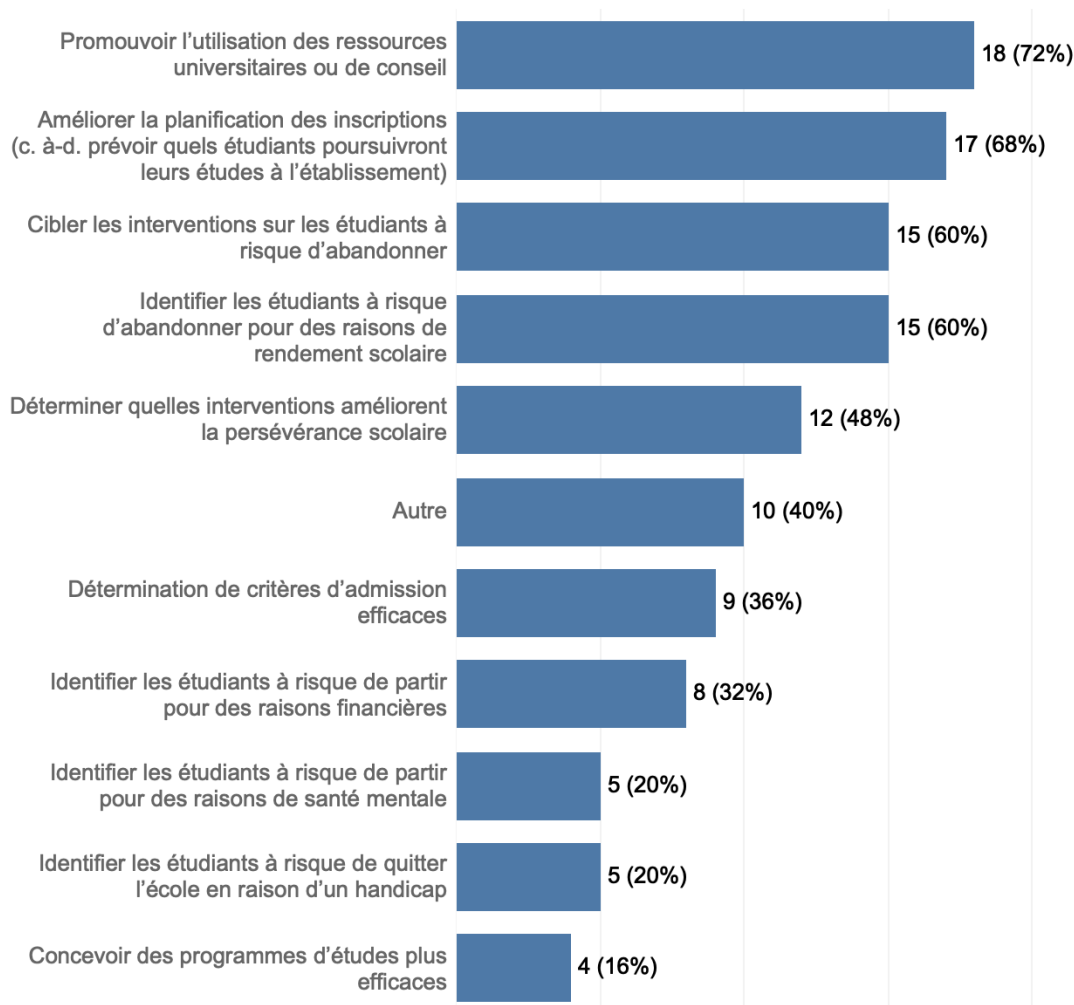
persévérance scolaire comme la promotion des services de soutien disponibles, les conseils, le mentorat ou les outils d'autoévaluation.

Types d'utilisation

À la question 5, on a demandé aux répondants comment la modélisation prédictive était utilisée dans leur établissement; les résultats sont présentés à la figure 12. Il s'agit d'une question à choix multiple et les options qui ont obtenu la majorité des 25 réponses possibles comprennent « identifier les étudiants à risque d'abandonner pour des raisons de rendement scolaire » avec 15 (60 %), « cibler les interventions sur les étudiants à risque de quitter » avec 15 (60 %), « promouvoir l'utilisation des ressources scolaires et/ou des conseils » avec 18 (72 %), « déterminer quelles interventions améliorent la persévérance scolaire » avec 12 (48 %) et « améliorer la planification scolaire » avec 17 (68 %). Il y a un chevauchement important dans la nature de ces utilisations – par exemple, cibler les interventions et déterminer l'efficacité de ces interventions vont de pair.

Moins de répondants ont indiqué qu'ils utilisaient la modélisation prédictive pour identifier les étudiants à risque d'abandonner les études pour des raisons non scolaires, comme des raisons de santé mentale (cinq, 20 %), d'invalidité (cinq, 20 %) ou financières (huit, 32 %), pour déterminer des critères d'admission efficaces (neuf, 36 %) ou pour concevoir des programmes plus efficaces (quatre, 16 %). Dans la catégorie « Autre », un répondant (4 %) a indiqué utiliser des modèles prédictifs pour l'établissement du calendrier des cours, le logement, l'évaluation du lien entre les écoles secondaires fréquentées et le rendement postsecondaire et l'établissement d'objectifs de rétention, tandis que trois répondants (11 %) ont indiqué qu'ils s'en servaient pour planifier les inscriptions, mais non pour retenir les étudiants, un répondant ayant indiqué regarder les taux d'obtention de diplôme plutôt que de rétention, un autre utilisant une optique budgétaire plutôt que de rétention et trois (11 %) ayant indiqué rechercher des groupes à risque sans identifier certains étudiants dans ces groupes.

Figure 12 : Réponses à la question 5, « Vous avez indiqué que votre modèle prédictif est utilisé à des fins de persévérance scolaire. Comment votre établissement utilise-t-il la modélisation prédictive? » (Sélectionnez toute réponse qui s'applique)

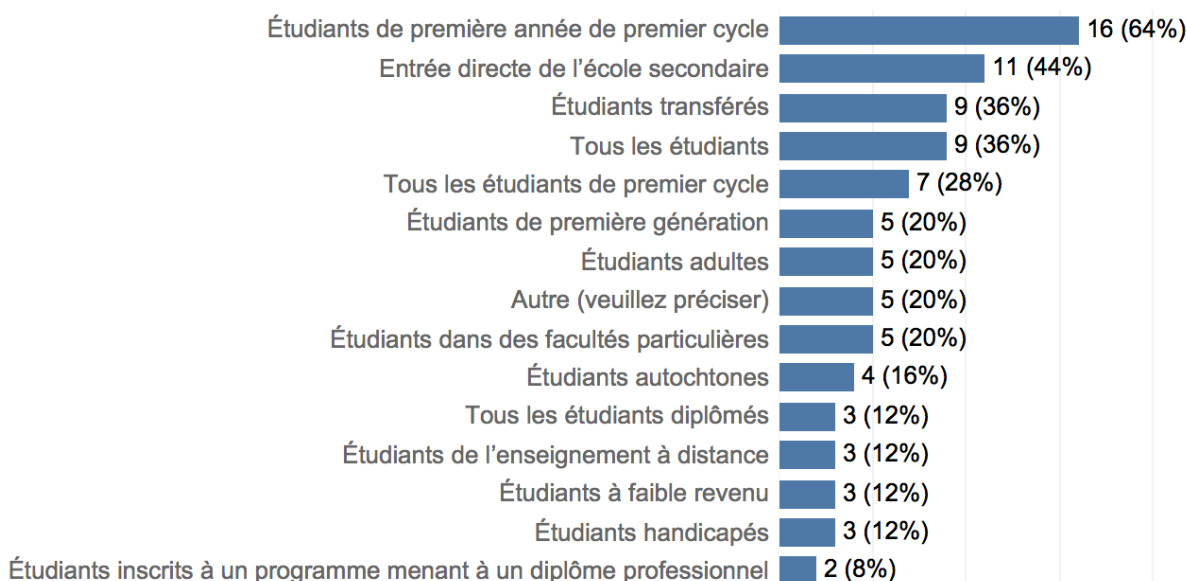


Populations étudiantes

La plupart des modèles prédictifs sont utilisés pour les étudiants de première année du premier cycle, avec 16 (57 %) des répondants ayant choisi cette réponse. Comme le montre la figure 13, c'est la seule réponse choisie par une majorité de répondants. Parmi les autres réponses fréquemment choisies, mentionnons les étudiants qui arrivent directement de l'école secondaire (11, 39 %), les étudiants transférés (neuf, 32 %), tous les étudiants (neuf, 32 %) et tous les étudiants de premier cycle (sept, 25 %). Des neuf répondants qui ont choisi « tous les étudiants », cinq ont également choisi au moins une catégorie supplémentaire, l'un d'eux ayant choisi toutes les options sauf « autre »; il en va de même pour « tous les étudiants de premier

cycle », la plupart (neuf sur dix) ayant également choisi une autre catégorie. Dans la catégorie « autre », deux répondants ont indiqué qu'ils utilisent la modélisation prédictive pour les étudiants en résidence, les autres répondants mentionnant les étudiants boursiers (un répondant), les étudiants de première année de maîtrise (un répondant) et les étudiants dans des facultés particulières. Lorsqu'on leur a demandé de donner plus de détails sur les « étudiants de facultés particulières », deux des cinq répondants ont indiqué qu'ils utilisaient la modélisation prédictive pour toutes les facultés et tous les programmes, mais qu'ils incluait la faculté ou le programme comme modèle d'entrée de la variable. D'autres répondants ont indiqué qu'ils utilisaient la modélisation prédictive pour les programmes d'entrée directe (un répondant) et les programmes réguliers (un répondant), et un répondant a mentionné expressément les facultés de droit et de génie.

Figure 13 : Réponses à la question 13, « Pour quelles populations étudiantes votre établissement utilise-t-il des modèles prédictifs pour la persévérance scolaire? » (Sélectionnez toute réponse qui s'applique)



Lorsqu'on leur a demandé dans quelle mesure leur modélisation prédictive était exacte, en matière de pourcentage de prédictions exactes relativement aux étudiants, dans la moitié des cas (36 sur 72), les répondants n'étaient pas certains de l'exactitude du système; parmi ceux qui ont répondu d'une façon quelconque, 27 sur 72 (38 %) ont indiqué 70 % ou plus. Voir le tableau 2 pour plus de détails et une ventilation par groupe d'étudiants.

Tableau 2 : Réponses à la question 19, « Veuillez indiquer quel pourcentage de prédiction par rapport aux étudiants est exact dans votre modélisation : »

	0-49 %	50-59 %	60-69 %	70-79 %	80-89 %	90-100 %	Incertain	Toutes les réponses
Étudiants de première année de premier cycle	1	1			2	2	8	14
Entrée directe de l'école secondaire	1	1	1		2	1	3	9
Tous les étudiants				4		2	2	8
Étudiants transférés	1				2	1	3	7
Tous les étudiants de premier cycle	1			1	1	3		6
Étudiants adultes				1	1		3	5
Étudiants autochtones		1			1		2	4
Étudiants de première génération				1			3	4
Autre (tel qu'identifié à la Q13)				1			3	4
Étudiants dans des facultés particulières							3	3
Étudiants à faible revenu							2	2
Étudiants de l'enseignement à distance					1		1	2
Étudiants handicapés		1					1	2
Tous les étudiants de cycle supérieur							1	1
Étudiants inscrits à un programme menant à un diplôme professionnel							1	1
Toutes les réponses	4	4	1	8	10	9	36	72

Données utilisées

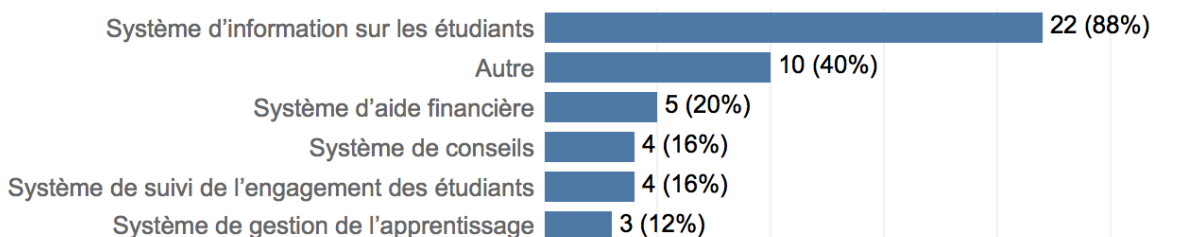
La source d'information la plus courante pour les systèmes de modélisation prédictive est le SIE, 22 répondants (88 %) ayant indiqué que leur système utilisait certaines informations du SIE (voir Figure 14). Parmi les autres options particulières, toutes ont été utilisées par entre trois (12 %) et cinq (20 %) répondants. Les systèmes de la catégorie « autres » comprenaient les systèmes de sondage et de questionnaire (trois), les systèmes de gestion des relations avec la clientèle étudiante éventuelle (deux), un système d'alerte précoce (un), les données conservées à l'extérieur d'un système particulier, comme les feuilles de calcul Excel (un), un SIE sans nom (un; non mentionné dans les 22 systèmes informatiques de

renseignements sur les étudiants) et un répondant qui ne donnait pas de détails. Un répondant n'a fourni aucune réponse.

La combinaison de réponses la plus courante était le système d'information sur les étudiants et autres (cinq, 20 %); aucune autre combinaison n'a obtenu plus d'une réponse. Les détails des combinaisons de réponses sont présentés à la figure 15.

Parmi les systèmes informatiques de renseignements sur les étudiants représentés, Ellucian's Banner a obtenu neuf réponses (41 %) et Oracle's PeopleSoft Campus Solutions trois (14 %), tandis que deux répondants (9 %) ont indiqué un système maison. D'autres réponses comprenaient « Ellucian »(qui pouvait être soit Banner, soit le produit Colleague d'Ellucian), « Oracle » (qui pouvait être PeopleSoft ou le système étudiant Oracle), tandis que Skytech et CrossRoad ont été mentionnés une fois chacun. Un répondant ne savait pas, tandis que trois n'ont fourni aucune information sur le produit. Les systèmes de gestion de l'apprentissage mentionnés comprennent Blackboard, Desire2Learn et Moodle; les systèmes de conseil étaient End2End et deux systèmes locaux; le système d'aide financière était Navison (d'autres ne connaissaient pas leur SIE ou n'en avaient pas tiré des données); et les systèmes de participation des étudiants comprenaient ezRecruit, CampusLabs et des systèmes locaux.

Figure 14 : Réponses à la question 11, « De quels systèmes vos modèles prédictifs de persévérance scolaire recueillent-ils l'information? » (Sélectionnez toute réponse qui s'applique)



La majorité des répondants ont indiqué qu'ils utilisaient les données démographiques sur les étudiants (20 répondants, 80 %), l'emplacement (14 répondants, 56 %) et les antécédents scolaires de leur établissement (17 répondants, 68 %) ou de l'école secondaire (16 répondants, 64 %) dans leurs modèles prédictifs (voir figure 16). Le nombre d'interactions du SGA était faible (trois répondants, 12 %), ce qui correspond aux réponses de la question 11 du SGA. Les questionnaires d'autoévaluation (huit répondants, 32 %) avaient tendance à être élaborés localement (quatre des huit réponses, 50 % des questionnaires d'autoévaluation), bien que le Consortium canadien de recherche sur les étudiants universitaires (deux des huit réponses, 25 % des questionnaires d'autoévaluation) et le National Survey of Student Engagement (deux des huit réponses, 25 % des questionnaires d'autoévaluation) aient été tous les deux mentionnés en tant que sondages externes. Parmi les tests normalisés (quatre, 16 %), le test d'habileté scolaire et l'ACT ont été inclus dans une réponse, tandis que deux réponses comprenaient des tests de langue anglaise, une réponse comprenait des examens de placement en anglais et en mathématiques élaborés localement et une réponse indiquait le law school admission test (LSAT). Les autres renseignements (huit, 32 %) comprenaient des renseignements sur l'admission et la demande, comme le moment où un étudiant a présenté sa demande et les programmes

auxquels il s'est inscrit (quatre réponses sur huit, 50 % des répondants indiquant d'autres renseignements); sa situation d'emploi (une réponse sur huit, 13 % des répondants indiquant d'autres renseignements); des renseignements financiers, notamment sur son aide financière (quatre réponses sur huit, 50 % des répondants indiquant d'autres renseignements); le lieu de résidence sur campus (une réponse sur huit, 13 % des répondants indiquant d'autres renseignements); l'examen des étudiants/facultés (une réponse sur huit, 13 % des répondants indiquant d'autres renseignements).

Figure 15 : Combinaison de réponses à la question 11, « De quels systèmes vos modèles prédictifs de persévérance scolaire recueillent-ils l'information? »

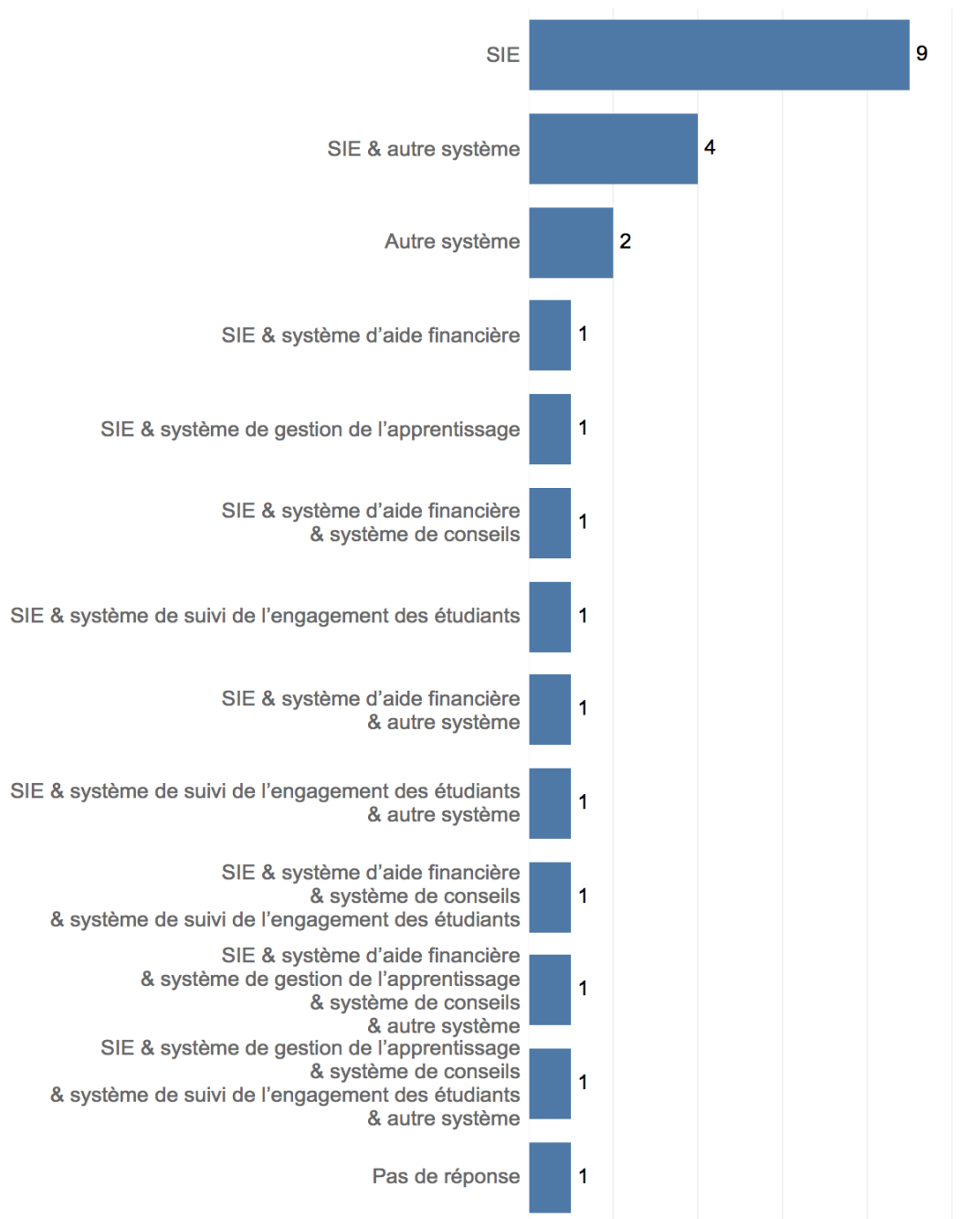
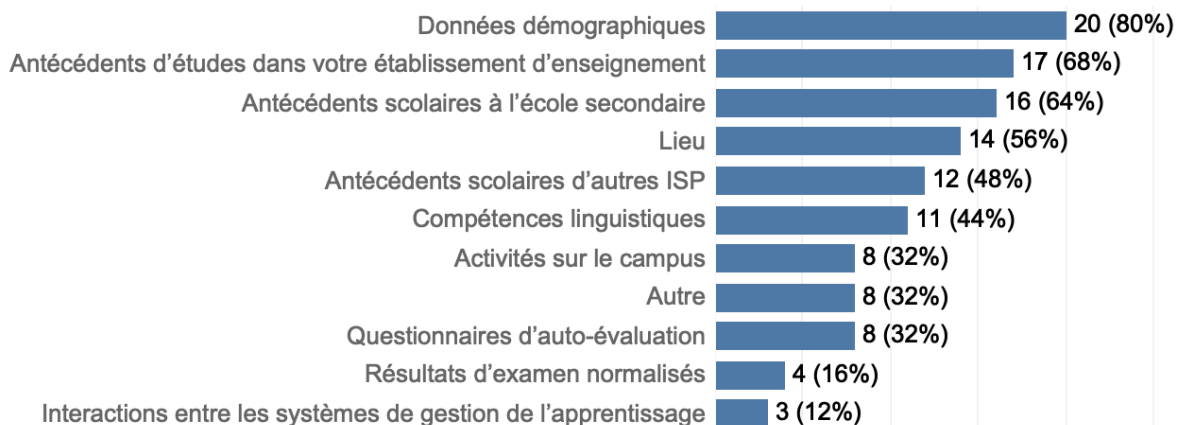


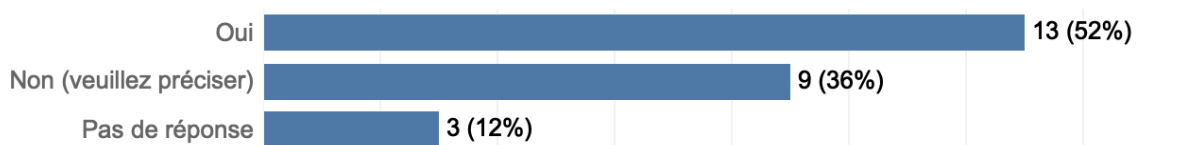
Figure 16 : Réponses à la question 12, « Quels types d’information vos modèles prédictifs utilisent-ils pour la persévérance scolaire? » (Sélectionnez toute réponse qui s’applique)



Interventions

Lorsqu'on leur a demandé s'ils utilisaient leur modèle prédictif pour éclairer des interventions spécifiques de persévérance scolaire, 13 répondants (52 %) ont répondu par l'affirmative, tandis que neuf répondants (36 %) ont répondu par la négative; trois répondants (12 %) n'ont pas répondu à la question (voir la figure 17). Parmi les répondants du groupe ayant répondu « non », trois ont indiqué qu'ils espéraient ou prévoyaient le faire à un moment donné, mais que leurs modèles étaient encore en cours d'essai et d'amélioration, tandis que deux répondants ont indiqué que les modèles n'étaient pas utilisés à ces fins, deux autres ont indiqué que l'information était fournie aux étudiants pour déterminer comment procéder, et un répondant a indiqué avoir fourni des exemples et des suggestions aux secteurs qui feraient les interventions, mais qui ne les avaient encore jamais utilisés.

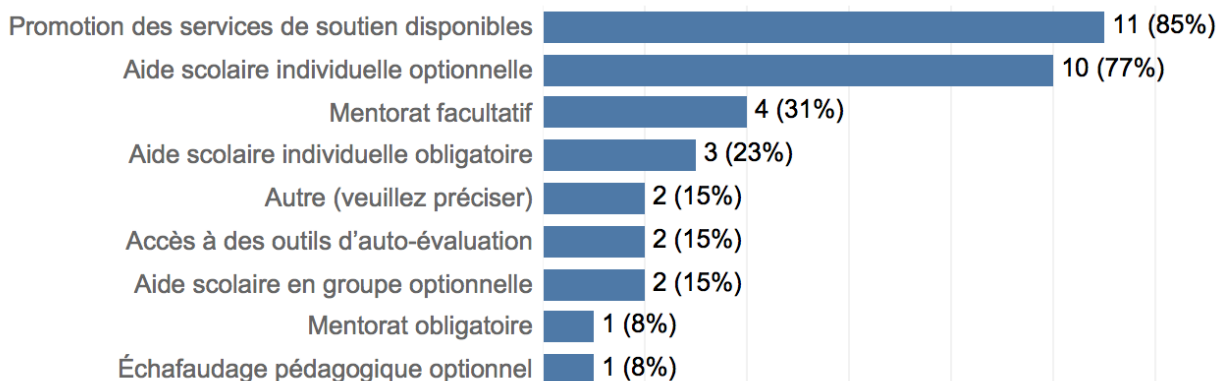
Figure 17 : Réponses à la question 20, « Utilisez-vous actuellement la modélisation prédictive pour orienter des interventions spécifiques en matière de persévérance scolaire? » (Sélectionnez une réponse)



Une partie de l'intention de la question 20 était que les questions 21 à 25 seraient uniquement posées aux répondants s'ils répondaient « oui ». Toutefois, cette logique n'a pas été appliquée à tort dans le sondage final. Les résultats ci-dessous pour les questions 21 à 25 ne sont fournis que pour les 13 répondants qui ont indiqué qu'ils utilisent la modélisation prédictive pour orienter des interventions spécifiques.

La question 21 demandait quelles interventions étaient utilisées parallèlement à la modélisation prédictive (voir la figure 18). Les groupes de réponses les plus importants étaient la promotion des services de soutien disponibles (11, 85 % des 13 répondants qui ont répondu « oui » à la question 20) et de l'aide scolaire individuelle optionnelle (10, 77 %). Les autres options ont sensiblement moins de réponses, avec seulement un petit nombre de réponses positives au mentorat optionnel (quatre, 31 %), à l'aide scolaire individuelle obligatoire (trois, 23 %), à l'aide scolaire de groupe optionnel (deux, 15 %), à l'échafaudage pédagogique optionnel (un, 8 %), à l'accès aux outils d'autoévaluation (deux, 15 %), aux autres (deux, 15 %) et au mentorat obligatoire (un, 8 %). Parmi ceux qui ont répondu « autre », un répondant a indiqué qu'ils avaient un programme de tutorat par les pairs, tandis qu'un autre a indiqué qu'ils prenaient des décisions en matière de programmation et de publicité fondées sur la modélisation prédictive à un niveau global, mais non au niveau d'un étudiant en particulier. Une autre réponse à « autre » était « alerte précoce », mais aucune autre information sur l'intervention spécifique associée à l'alerte précoce n'a été fournie.

Figure 18 : Réponses à la question 21, « Veuillez indiquer les différents types d'interventions que votre établissement utilise en fonction de vos modèles prédictifs ». (Sélectionnez toute réponse qui s'applique)



De nombreux participants aux entrevues/questionnaires ont fait état de nouvelles interventions liées à la modélisation prédictive. Certaines interventions étaient en place pendant le processus d'admission, où le type et le contenu des communications avec les demandeurs ont changé en raison de l'information provenant des modèles, tandis qu'un autre participant a signalé des changements dans les sources des demandeurs. D'autres interventions ont été utilisées auprès des étudiants pendant qu'ils fréquentaient l'établissement, comme l'amélioration des services d'approche auprès des étudiants, tandis qu'un participant a signalé un changement radical dans le modèle de prestation de conseils de l'établissement. Les participants ont également fait remarquer que les interventions ne sont pas toutes individuelles, un participant ayant mentionné expressément l'utilisation de modèles prédictifs pour convaincre la haute

direction d'augmenter le nombre de sections de cours offertes; cette augmentation a entraîné une amélioration des taux de rétention puisque les étudiants pouvaient obtenir davantage des cours qu'ils voulaient.

On a également demandé aux participants ce qui les avait aidés à passer d'une modélisation prédictive d'un outil plus théorique à un contexte appliqué. Bien qu'un participant ait fait état de compétences techniques en analyse de données et en gestion de données, la plupart des réponses portaient sur l'obtention d'un soutien de haut niveau et la capacité de générer des gains précoces à l'aide de modèles prédictifs. Il a été noté que le fait d'avoir des gains précoces permettait aux intervenants d'avoir confiance dans les décisions prises en se basant sur les modèles prédictifs.

Possibilités et défis de la modélisation prédictive

Comme les modèles prédictifs ne sont pas utilisés depuis de nombreuses années dans la plupart des établissements, les répondants signalent qu'il est souvent trop tôt pour dire quel en est l'effet. Bien que, dans certains cas, les établissements signalent qu'ils voient un effet positif dans diverses mesures de la réussite des étudiants, comme les taux de rétention et de persévérance, les taux de réussite scolaire, les taux d'obtention du diplôme et le rendement des étudiants sur des mesures comme la moyenne pondérée cumulative. Un répondant a fait remarquer qu'il y a trop de variables pour déterminer l'effet spécifique de la modélisation prédictive et des interventions. L'un des succès notables des modèles prédictifs est qu'ils ont provoqué un changement dans la culture et le type de conversations qui ont lieu dans certains établissements et qu'ils fournissent des preuves du changement, ainsi que l'introduction de nouveaux soutiens et/ou des changements dans la façon dont les soutiens scolaires sont favorisés. Grâce aux modèles prédictifs, des changements ont été apportés à l'aide scolaire, aux politiques de persévérance scolaire, à la conception de programmes d'études et à l'échafaudage pédagogique, aux programmes d'orientation et aux mesures de transition pour la première année.

Les défis les plus souvent cités par les répondants à l'égard de l'utilisation de la modélisation prédictive comprennent les problèmes liés aux données (p. ex. les efforts requis pour recueillir et nettoyer les données, l'exactitude et le calendrier des données), les taux de réponse au sondage, le manque de ressources (temps, compétences) et l'acceptation dans les établissements. Interrogés sur les conseils qu'ils offriraient aux établissements qui ne s'étaient pas encore lancés dans la modélisation prédictive, les répondants ont souligné l'importance de faire participer les intervenants, de susciter leur adhésion et d'être réalistes quant à l'étendue des ressources (p. ex. connaissances, compétences et temps) requises pour élaborer et maintenir des modèles prédictifs. Les types de nouvelles données que les établissements aimeraient intégrer à leurs modèles comprennent le rendement scolaire, des données démographiques supplémentaires, des données de sondage auprès des étudiants, des données de CRM, l'utilisation de divers services sur les campus ou l'implication dans des incidents sur les campus, des données financières et des renseignements provenant d'un SGA.

Dans de nombreux cas, les répondants estimaient que le modèle prédictif utilisé répondait à leurs besoins. Lorsqu'on estimait que le système ne répondait pas aux besoins, les raisons invoquées comprenaient le fait que la modélisation prédictive n'en était qu'à ses débuts, le manque d'utilisation de la modélisation au sein

de leur établissement et le manque de temps et/ou de ressources. La plupart des modèles sont examinés régulièrement ou au besoin par le personnel interne, parfois avec l'aide du corps professoral ou des fournisseurs de systèmes internes.

Dans un environnement où la prise de décisions et l'affectation des ressources fondées sur des données probantes sont de plus en plus courantes, il est clair que la modélisation prédictive pourrait jouer un rôle important dans les années à venir, à mesure que les modèles se perfectionneront et se développeront avec le temps.

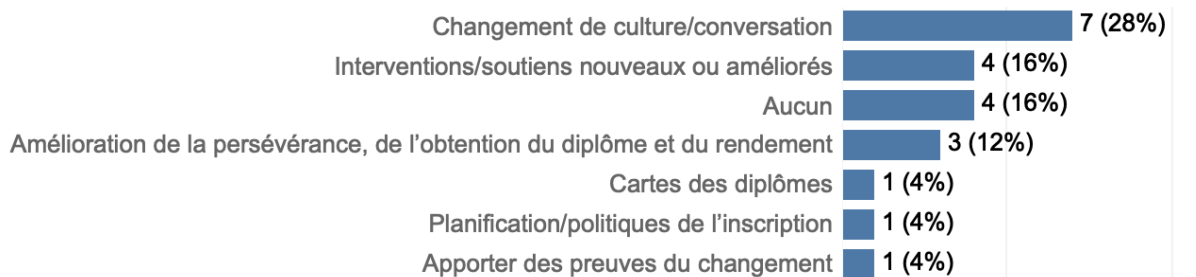
Effets de l'utilisation de la modélisation prédictive

Les questions de 22 à 25 demandaient quel était l'effet de la modélisation prédictive sur les diverses mesures de la réussite des étudiants, que les établissements peuvent surveiller comme les taux de rétention et de persévérance, les taux de réussite scolaire, les taux d'obtention du diplôme et le rendement des étudiants sur des mesures comme la moyenne pondérée cumulative. Dans la majorité des cas, les répondants ont indiqué qu'il était trop tôt pour voir s'il y avait un effet (entre trois et sept réponses; de 23 à 54 %), qu'ils ne savaient pas ou qu'il n'y avait aucun effet (entre deux et six réponses; de 16 à 46 %). Entre un et cinq répondants (de 8 à 38 %) ont indiqué qu'il y avait un effet positif. Un répondant qui a indiqué une incidence positive a fait état d'une augmentation d'environ 5 % des taux de rétention, un autre d'une augmentation globale de 2 %, certains groupes, comme les Autochtones, ayant enregistré des gains plus élevés, tandis que les autres n'ont pu fournir de chiffres précis. Un répondant indiquant une combinaison d'impacts positifs et négatifs a fait remarquer qu'aucune analyse systématique n'avait été effectuée, mais que, de façon empirique, les modèles avaient permis de repérer certains étudiants qui auraient été oubliés par d'autres méthodes ainsi que certains étudiants qui n'avaient pas besoin d'interventions. Parmi les répondants qui ont indiqué un effet positif sur les taux d'obtention ou d'achèvement du diplôme, un répondant a indiqué une amélioration de 2 %, un autre une amélioration de 5 à 10 % chaque année sur plusieurs années, et un autre a indiqué qu'il y avait trop de variables pour identifier l'effet spécifique de la modélisation prédictive et des interventions. Les résultats plus détaillés de cette série de questions se trouvent à l'annexe C : Les questions 22 à 25 et les figures 32 à 35 portent respectivement sur la persévérance scolaire, les résultats scolaires, les taux d'obtention d'un diplôme et d'achèvement et le rendement, comme la moyenne pondérée cumulative.

Lorsqu'on leur a demandé quels étaient les plus grands succès dans l'utilisation de la modélisation prédictive pour la persévérance scolaire, le plus grand groupe de répondants (sept, 28 %) ont indiqué que l'utilisation de la modélisation prédictive avait provoqué un changement dans la culture et la conversation autour de l'établissement; les répondants ont parlé de « concentrer la conversation sur la persévérance scolaire » et comment la modélisation prédictive « a fait parler de la réussite des étudiants »; un répondant (4 %) a également écrit à ce sujet, notamment en matière d'interventions, à l'appui des changements dans leur établissement. Quatre répondants (16 %) ont indiqué que la création ou l'amélioration des interventions et des mesures de soutien des étudiants avait été un grand succès, tandis que trois répondants (12 %) ont constaté des améliorations dans l'obtention de diplômes, la rétention et le rendement des étudiants. Les cartes des diplômes – où le parcours qu'un étudiant doit suivre dans le programme d'études d'un

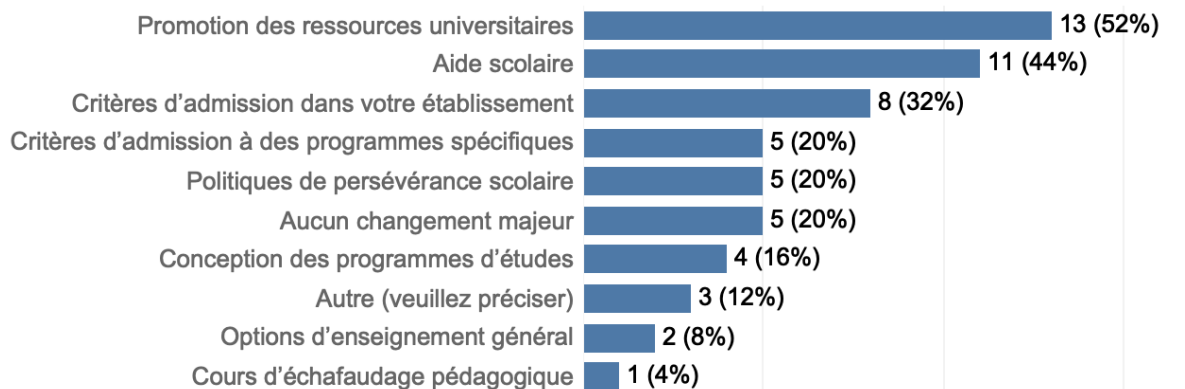
programme est présenté – et l'utilisation de données probantes dans la planification et les politiques d'inscription ont été mentionnées par un répondant (4 %) chacun. En plus des quatre répondants (16 %) qui ont indiqué une version quelconque de « sans réussite », huit répondants (32 %) n'ont pas répondu à cette question (voir la figure 19).

Figure 19 : Réponses à la question 27, « Quelles sont les plus grandes réussites que vous avez eues en matière de modélisation de la persévérance scolaire? » (Qualitativement codé)



La question 32 demandait quels changements s'étaient produits dans les établissements avec l'utilisation de la modélisation prédictive. Comme le montre la figure 20, 13 répondants (52 %) utilisant la modélisation prédictive ont indiqué que des changements avaient été apportés à la façon dont les ressources universitaires étaient promues, tandis que 11 répondants (44 %) ont apporté des changements aux services d'aide scolaire dans leur établissement. Un petit nombre de répondants ont mentionné des changements aux critères d'admission au niveau de l'établissement (huit, 32 %) ou du programme (cinq, 20 %), des changements aux politiques de persévérance scolaire (cinq, 20 %), à la conception du programme (quatre, 16 %), aux options de formation générale (deux, 8 %) et aux cours sur le soutien scolaire (un, 4 %). Dans le groupe « autre » (trois, 12 %), deux répondants ont identifié des changements à l'orientation et à la transition des étudiants dans l'établissement et l'autre a identifié des changements dans la façon dont les décisions de réussite ou d'échec sont prises pour certains étudiants, sans toutefois fournir de renseignements supplémentaires.

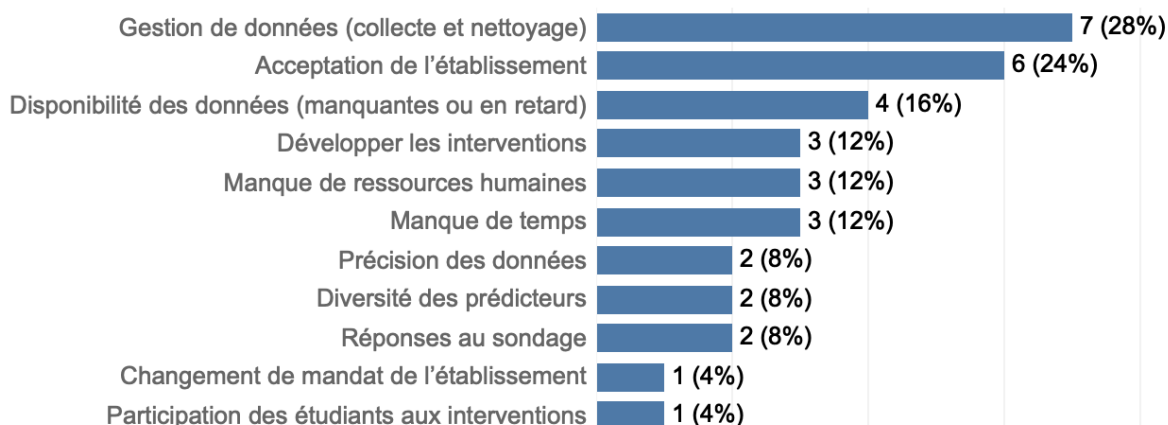
Figure 20 : Réponses à la question 32, « L'utilisation de la modélisation prédictive dans votre établissement a-t-elle entraîné des changements importants dans les domaines suivants? » (Sélectionnez toute réponse qui s'applique)



Défis à relever

Interrogés sur les difficultés qu'ils avaient rencontrées dans l'utilisation de modèles prédictifs pour la persévérance scolaire, bon nombre des répondants au sondage ont parlé de problèmes liés aux données, comme l'effort requis pour « la gestion » (collecte et nettoyage) des données (sept, 28 %), le traitement des données manquantes ou qui mettent beaucoup de temps à être disponibles (quatre, 16 %), la diversité des éléments qui pourraient être utiles comme prédicteurs (deux, 8 %), la précision des données qui sont accessibles (deux, 8 %) et l'obligation de remplir le questionnaire (deux, 8 %). L'acceptation (six, 24 %) était un autre défi souvent cité, les répondants indiquant qu'il était difficile de faire accepter et adopter les modèles prédictifs dans la communauté. Le manque de ressources a été cité comme un défi à la fois pour le temps (trois, 12 %) et les personnes qualifiées (trois, 12 %); l'élaboration d'interventions (trois, 12 %) et le fait d'amener les étudiants à participer aux interventions (un, 4 %) ont également été mentionnés. Six répondants n'ont pas répondu à cette question.

Figure 21 : Codage qualitatif des réponses à la question 26, « Quelles sont les plus grandes réussites que vous avez eues en matière de modélisation de la persévérance scolaire? »

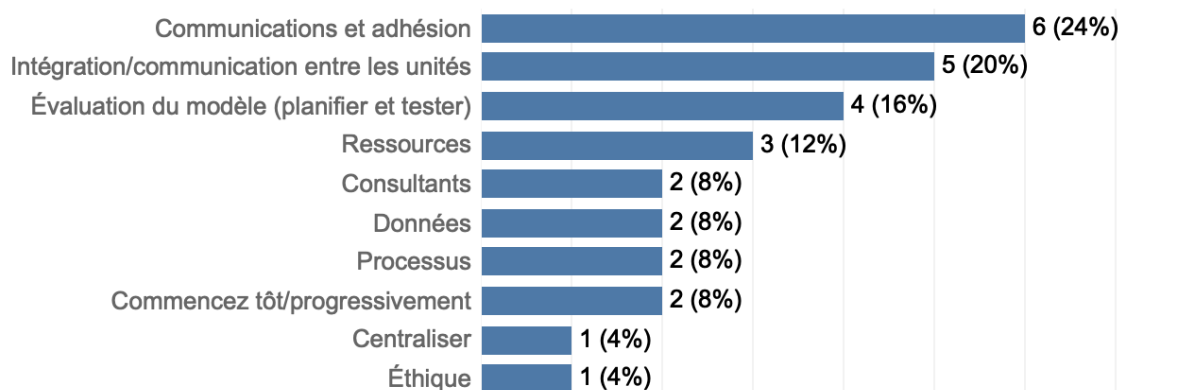


Les réponses aux entrevues et au questionnaire concernant les défis à relever reflétaient en grande partie les réponses à une question semblable sur les sondages – les problèmes liés à la collecte et au nettoyage des données, à l'affectation appropriée des ressources et au niveau de connaissances ont de nouveau été relevés. Toutefois, deux nouveaux éléments sont également apparus, le premier étant la difficulté de traiter de grands ensembles de données et les exigences en matière de calcul des modèles prédictifs, et la façon dont cela peut être un fardeau pour l'infrastructure de technologie de l'information existante. L'autre nouveauté est la question de l'« intuitivité » de la modélisation et le fait qu'il n'est pas toujours simple ou facile pour les concepteurs de modèles d'expliquer à ceux qui doivent utiliser le modèle comment comprendre les prédicteurs sélectionnés par le modèle.

Interrogés sur les conseils qu'ils offriraient aux établissements qui ne s'étaient pas encore lancés dans la modélisation prédictive de la persévérance scolaire, la majorité des répondants (six, 24 %) ont suggéré de communiquer avec les intervenants et de susciter leur adhésion, en formulant des commentaires tels que « expliquez clairement pourquoi », « il faut de la patience » et « il est important d'acquérir des connaissances ». Dans le même ordre d'idées, cinq répondants (20 %) ont parlé de l'intégration et de la communication entre les unités, en particulier de l'endroit où les données sont générées jusqu'à l'endroit où la modélisation est effectuée, et finalement aux groupes qui font les interventions et autres programmes. Quatre répondants (16 %) ont mentionné avoir un plan pour évaluer régulièrement la modélisation prédictive et trois répondants (12 %) ont mentionné avoir les ressources nécessaires. D'autres ont suggéré de commencer tôt (deux, 8 %), d'étudier et de connaître les données (deux, 8 %), d'avoir un processus d'élaboration, de mise à l'essai et de discussion avec les intervenants au sujet du modèle (deux, 8 %), de prêter attention aux dilemmes éthiques inhérents à la modélisation prédictive et d'en parler (un, 4 %), et de centraliser l'élaboration du modèle (un, 4 %). Le recours à des consultants externes a été évoqué dans deux réponses, mais de manière diamétralement opposée : Un répondant a indiqué que « les consultants

externes peuvent fournir un soutien important pour l'élaboration du processus de création et de socialisation du modèle » tandis qu'un autre a simplement dit « ne pas faire confiance aux consultants ».

Figure 22 : Codage qualitatif des réponses à la question 28, « Quels conseils donneriez-vous à un établissement qui cherche à mettre en œuvre un modèle prédictif pour la persévérance scolaire? »



Lorsqu'on leur a demandé ce qu'ils auraient aimé savoir lorsqu'ils ont commencé à utiliser la modélisation prédictive, les participants ont répondu presque unanimement qu'ils possédaient les connaissances et les ressources nécessaires pour maintenir et améliorer le modèle. Dans certains cas, les préoccupations portaient sur le ressourcement technique, comme le choix du bon type de logiciel ou l'assurance que le matériel disponible est adéquat pour la tâche, tandis que dans d'autres cas, il s'agissait du ressourcement du personnel et de la compréhension non seulement du temps requis, mais aussi des compétences et connaissances requises. Deux participants ont spécifiquement mentionné le fait de commencer par des modèles externes et de ne pas bien comprendre les besoins en ressources lorsqu'il s'agit de placer ces modèles sous le contrôle des établissements.

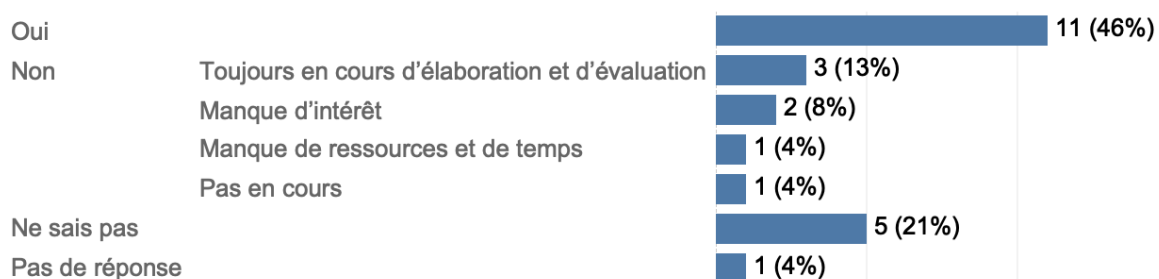
Plus de la moitié des répondants (13, 52 %) ont indiqué qu'ils aimeraient intégrer certains types de données qui ne sont pas actuellement utilisés dans leur modélisation prédictive. Les types de données qui pourraient être utilisées comprennent le rendement scolaire, soit au niveau secondaire ou postsecondaire (quatre), des renseignements démographiques supplémentaires (quatre), des données d'enquête auprès des étudiants (trois), d'autres données sur les établissements, comme les renseignements sur le CRM ou la participation aux incidents sur les campus (trois) et des données financières (deux).

Évaluer si les systèmes de modélisation prédictive répondent aux besoins des établissements

De nombreux répondants au sondage (11, 46 %) estiment que leur système et leurs méthodes de modélisation prédictive actuels répondent bien aux besoins de leur établissement (voir la figure 23). Cinq autres répondants (21 %) ne savaient pas si leur système répondait bien à leurs besoins, tandis que sept

répondants (29 %) ont indiqué que ce n'était pas le cas. Parmi ceux qui estimaient que leur système et leurs méthodes ne fonctionnaient pas bien, trois répondants (13 %) ont indiqué que leur modélisation prédictive en était encore à ses débuts, deux d'entre eux invoquant un manque d'utilisation de la modélisation au sein de leur établissement, un autre qu'un manque de ressources et de temps les retenait, et un répondant que leur modélisation prédictive ne constituait pas encore un processus continu.

Figure 23 : Réponses à la question 18, « Trouvez-vous que votre système et vos méthodes de modélisation prédictive actuels répondent à vos besoins? » (Sélectionnez une réponse)



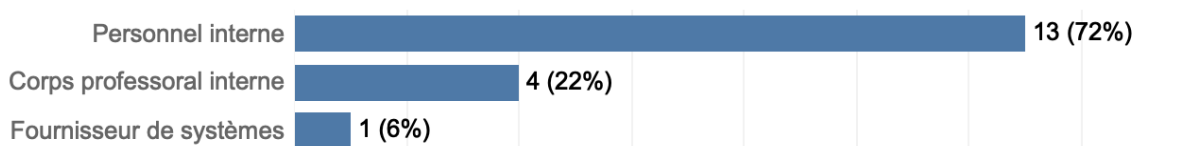
Lorsqu'on leur a demandé si leur système faisait l'objet d'examen réguliers, les réponses étaient également partagées entre le fait que leur modélisation n'avait pas été examinée, qu'elle faisait l'objet d'un examen annuel et qu'elle était examinée régulièrement, mais non annuellement (voir la figure 24).

Figure 24 : Réponses à la question 30, « Votre système de modélisation prédictive a-t-il été examiné depuis sa mise en œuvre initiale pour voir s'il est possible de l'améliorer? » (Sélectionnez une réponse)



Lorsque les répondants ont indiqué que leur modèle avait été examiné régulièrement ou annuellement, la question suivante leur demandait qui avait participé au processus d'examen du modèle. Comme le montre la figure 25, le personnel interne a été l'option la plus choisie avec 13 réponses (72 % des 18 répondants ont posé cette question), et un petit nombre a indiqué qu'il faisait appel à des chargés de cours internes (quatre, 22 %) ou au fournisseur du système (un, 6 %).

Figure 25 : Réponses à la question 31, « Cet examen du modèle est-il un processus interne ou travaillez-vous avec une organisation externe? » (Sélectionnez toute réponse qui s'applique)



La question visant à savoir comment l'utilisation de la modélisation prédictive a évolué depuis la mise en œuvre initiale a été interprétée dans les entrevues et les questionnaires sous deux angles différents : Certaines réponses portaient sur la façon dont la modélisation elle-même avait changé, tandis que d'autres portaient sur la façon dont l'utilisation de la modélisation prédictive avait changé le contexte de l'établissement. Parmi les changements apportés à la modélisation, les réponses comprenaient des améliorations aux données utilisées, comme l'augmentation du volume de données, du nombre de prédicteurs ou de l'exactitude (six, 67 %) et l'augmentation de la fréquence avec laquelle la modélisation est effectuée (deux, 22 %). Parmi les changements par rapport à la modélisation citée, mentionnons une augmentation de la tolérance au risque fondée sur la capacité de modéliser les changements avant la mise en œuvre (un, 11 %) et l'utilisation de la modélisation prédictive dans les décisions de résidence (un, 11 %) et dans les interventions ciblées (un, 11 %). De plus, un participant (11 %) a répondu que l'utilisation de la modélisation prédictive avait fourni plus d'information et de données probantes au comité de gestion stratégique des inscriptions de l'établissement et qu'elle avait aidé à ouvrir la voie à un modèle budgétaire qui était plus étroitement lié aux inscriptions. (Voir la figure 26.)

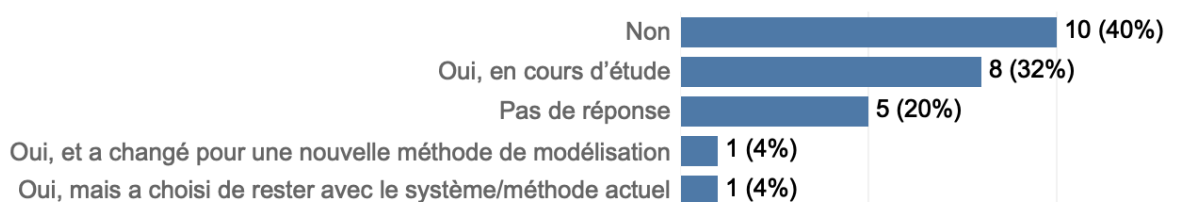
Figure 26 : Indication de l'évolution de la modélisation prédictive des participants à la phase deux



On a également demandé aux participants comment ils voyaient l'évolution de l'utilisation de la modélisation prédictive par leur établissement au cours des cinq prochaines années. Les réponses ont renforcé les thèmes clés identifiés dans le sondage – apporter de nouvelles données dans le processus de modélisation et veiller à ce que les modèles soient continuellement évalués et améliorés. Les participants ont également indiqué qu'ils souhaitaient que leurs modèles prédictifs soient utilisés pour rendre les établissements plus centrés sur l'étudiant, l'un d'entre eux ayant spécifiquement parlé de l'utilisation de l'argent des établissements pour encourager les changements à cette fin.

La dernière grande question du sondage demandait aux répondants s'ils avaient étudié d'autres méthodes de modélisation prédictive par rapport à leurs méthodes actuelles. Comme le montre la figure 27, les plus grands groupes de répondants ont indiqué qu'ils n'avaient pas examiné les solutions de rechange (10, 40 %) ou qu'ils étaient en train de le faire (huit, 32 %), tandis qu'un répondant (4 %) avait examiné les solutions de rechange, mais avait choisi de conserver sa méthode actuelle. Un répondant (4 %) qui avait examiné le tout a finalement changé d'avis, et lorsqu'on lui a demandé d'en parler davantage, il a indiqué qu'il était passé d'un modèle simpliste à un modèle qui tenait compte en plus des taux de chômage locaux. Cinq autres répondants (20 %) n'ont pas répondu à cette question.

Figure 27 : Réponses à la question 33, « Avez-vous étudié d'autres méthodes de modélisation de la persévérance scolaire, comme l'utilisation d'un système différent ou de méthodes de prévision différentes? » (Sélectionnez une réponse)



Discussion

Cette recherche sur l'utilisation de la modélisation prédictive dans les établissements d'enseignement postsecondaire a révélé plusieurs thèmes généraux et jeté de la lumière sur les questions de recherche. Dans la discussion qui suit, nous commençons par examiner l'utilisation actuelle de la modélisation prédictive dans les établissements postsecondaires et les pratiques et principes qui s'y appliquent. Nous examinons ensuite comment la modélisation prédictive est utilisée de façon novatrice et mise à profit pour améliorer les décisions, ainsi que les possibilités et les défis liés à la modélisation prédictive identifiés par les répondants. Enfin, nous discutons des orientations futures de la recherche et des limites de cette recherche.

Utilisation de la modélisation prédictive dans les établissements postsecondaires

Sur les 70 réponses à notre sondage, 36 % des répondants ont indiqué qu'ils utilisaient la modélisation prédictive d'une façon ou d'une autre au sein de leur établissement et 39 % ont répondu qu'ils envisageaient ou envisageaient sérieusement de le faire dans un proche avenir. Des modèles prédictifs ont également été utilisés dans plus d'un contexte dans de nombreux cas, comme l'utilisation d'un modèle global à des fins d'inscription ou de planification budgétaire, et d'un modèle plus individuel pour les interventions auprès des étudiants à risque ou la promotion de l'utilisation des ressources de conseils. Les 68 % des répondants qui ont indiqué qu'ils utilisaient la modélisation prédictive à des fins de planification des inscriptions ont également indiqué qu'ils utilisaient la modélisation prédictive à des fins plus spécifiques

aux étudiants, et 44 % des participants aux entrevues et questionnaires utilisaient les deux types de modélisation.

Lorsqu'on leur a donné l'occasion d'indiquer pourquoi ils utilisaient la modélisation prédictive, de nombreux répondants ont indiqué une certaine pression sur l'établissement ou le milieu universitaire. Cette pression pourrait être d'ordre financier, à mesure que les budgets des établissements se resserrent ou que la démographie locale se déplace, ou encore, elle pourrait être une pression pour améliorer la réussite des étudiants. Bien qu'un certain nombre de répondants aient indiqué qu'ils n'avaient pas recours à la modélisation prédictive en raison de problèmes de données ou de ressources, deux répondants au sondage ont indiqué que l'adoption de la modélisation prédictive n'était pas nécessaire, car les taux de rétention et d'obtention du diplôme dans leur établissement étaient déjà élevés.

L'adoption de la modélisation prédictive a également augmenté au cours des dernières années. La moitié des répondants qui ont répondu à cette question dans le sondage (11 sur 25) ont indiqué que leurs modèles prédictifs avaient été mis en œuvre entre 2014 et 2017, et cinq autres ont indiqué qu'ils avaient un modèle prédictif en développement.

Ces données suggèrent que la modélisation prédictive dans l'enseignement postsecondaire canadien en est à ses débuts. Il est clair que les établissements qui utilisent la modélisation prédictive en sont encore à apprendre les avantages et les défis associés au processus, aux outils et aux résultats de celle-ci. Par conséquent, les établissements qui cherchent à adopter la modélisation prédictive sont moins susceptibles de trouver des pratiques exemplaires ou des outils qui peuvent répondre à leurs propres besoins. De plus, 56 % des répondants ont indiqué que la modélisation est un processus itératif qui fait l'objet d'un examen régulier ou annuel, ce qui donne à penser que les modèles continueront d'être perfectionnés au cours des prochaines années. Les recherches indiquent qu'il existe peu de ressources empiriques éclairées qui peuvent aider les établissements à établir des modèles prédictifs (Gašević et coll., 2016). Cette étude, jumelée à d'autres recherches et au partage interinstitutionnel, contribuera à cette base de connaissances, en particulier au Canada.

Pratiques et principes de la modélisation prédictive

Il y avait relativement peu de questions pour lesquelles la plupart ou la totalité des répondants étaient clairement d'accord, à deux exceptions près. La première consiste à s'entendre sur l'affectation appropriée des ressources nécessaires à la modélisation prédictive, tant en ce qui concerne les compétences des personnes chargées d'élaborer et d'utiliser des modèles prédictifs que le temps et les infrastructures techniques nécessaires pour les appuyer. La modélisation prédictive nécessite des ensembles d'outils logiciels spécialisés, une importante capacité de calcul et l'accès à des sources de données souvent diverses. La recherche sur le financement de la modélisation prédictive a été limitée jusqu'à présent, mais Gašević et coll. (2016) mentionnent l'importance d'un leadership éclairé et engagé comme l'un des principaux attributs d'une mise en œuvre réussie.

Le deuxième point d'accord concernait la sécurité des modèles prédictifs. L'accès aux résultats individuels prévus était très étroitement limité à ceux qui avaient besoin de savoir pour faire fonctionner les modèles,

évaluer les modèles ou coordonner ou exécuter des interventions. Bien que la plupart des répondants aient indiqué qu'il n'y avait pas de politiques précises en place concernant la modélisation prédictive, les personnes interrogées et les répondants au questionnaire ont parlé de l'importance de traiter les prédictions de façon confidentielle et comme des renseignements personnels sensibles en vertu des lois pertinentes sur la protection des renseignements personnels.

Gašević et coll. (2016) ont invité un groupe d'experts à compléter l'énoncé suivant : « *Pour que l'analyse de l'apprentissage ait un impact continu sur l'apprentissage et l'enseignement, il faudrait qu'elle...* » et à effectuer une analyse de regroupement sur les réponses. Ils ont déterminé six groupes : plateforme de données – normes et gouvernance (y compris des modèles de partage et une garantie de sécurité); utilisation des données – accessibles, transparentes, valides et fiables; compatibilité avec les valeurs/pratiques/systèmes existants; stratégie – vue d'ensemble de l'organisation; outils réalisables avec une base factuelle; soutien à l'autonomisation des étudiants. Ainsi, la modélisation passe de « quelque chose fait aux éducateurs et aux étudiants, [à] quelque chose fait avec eux en partenariat » (Gašević et coll., 2016, p.20).

Utilisations novatrices des modèles prédictifs

Les grandes lignes de la modélisation prédictive de la persévérance scolaire étaient assez semblables pour la plupart des répondants. Des utilisations similaires reposaient sur des techniques et des approches semblables. Toutefois, plusieurs réponses méritent d'être soulignées comme étant uniques d'une certaine façon, notamment l'utilisation de la modélisation prédictive pour les espaces dans les résidences étudiantes et la mise à la disposition directe des étudiants des résultats des modèles prédictifs.

Bien qu'un certain nombre de répondants aient indiqué qu'ils utilisaient les données sur la résidence comme données d'entrée dans leur modèle – c'est-à-dire si un étudiant vivait ou non dans une résidence sur le campus ou dans une résidence maintenue par un établissement – un seul établissement a indiqué qu'il utilisait un modèle prédictif pour aider à savoir qui se ferait offrir une place en résidence. Si un étudiant ne réussissait pas à terminer son premier semestre, il n'y aurait que peu ou pas d'occasion de combler cette place en résidence jusqu'à la prochaine demande de résidence et au prochain cycle d'offre, souvent dans plusieurs mois, ce qui entraîne non seulement une perte de revenu pour l'établissement, mais aussi une perte de capacité de fournir une place à quelqu'un qui pourrait en avoir besoin.

Deux répondants ont indiqué qu'au lieu de laisser les résultats de la modélisation prédictive entre les mains du corps professoral et du personnel de l'établissement, ils ont rendu l'information directement accessible aux étudiants. Ils ont couplé les prédictions avec de l'information qui a permis aux étudiants de déterminer leur propre cheminement en choisissant les mesures de soutien disponibles qui pourraient les intéresser. Comme aucun de ces établissements n'a choisi de participer à la deuxième phase de l'étude, il n'est pas possible de fournir des détails supplémentaires de ces établissements sur la façon dont ce processus fonctionne pour les étudiants.

Certains répondants ont indiqué que leur utilisation de la modélisation prédictive était liée à leur plan de gestion stratégique des effectifs de l'établissement. Un répondant en particulier a fait un pas de plus et a inclus son comité de gestion stratégique des inscriptions dans l'évaluation et l'examen annuels de sa modélisation prédictive, ainsi que dans un processus d'approbation explicite. Parmi les informations présentées au comité figurent les variables choisies pour le modèle prédictif ainsi que leur pondération relative. La discussion permet aux membres du comité de traiter des éléments qui devraient possiblement être pris en compte dans le modèle et de voir comment les prédicteurs changent d'une année à l'autre, tout en examinant le rendement passé.

Possibilités et défis des interventions

Il n'est pas simple d'utiliser la modélisation prédictive pour éclairer les interventions auprès des étudiants; seulement 57 % des répondants au sondage qui ont un modèle prédictif l'utilisent parallèlement à une intervention. Dans certains cas, il s'agissait simplement d'une question de temps – les répondants ont indiqué que le plan consistait à relier les deux, mais cela ne s'était pas encore produit – mais il y avait aussi des indications dans le sondage et dans les réponses aux entrevues et aux questionnaires qu'il était difficile d'amener les responsables des interventions à voir la valeur dans la modélisation. L'autre grand défi cité concerne le personnel possédant les compétences et le temps nécessaires pour rassembler les données, évaluer les prédicteurs et élaborer des modèles.

Interrogés sur les répercussions de la modélisation et des interventions prédictives, de nombreux répondants ont indiqué qu'il était trop tôt pour le savoir ou qu'ils n'avaient pas vu d'incidence sur les diverses mesures qui pourraient être utilisées, comme la rétention, le bon rendement scolaire, les taux d'obtention de diplôme ou les moyennes pondérées cumulatives. Un répondant a fait remarquer qu'il y a trop de variables pour déterminer l'effet spécifique de la modélisation prédictive et des interventions. Il pourrait s'agir d'une occasion d'utiliser des procédures de conception expérimentale appropriées, semblables à celles mentionnées dans l'analyse documentaire pour la deuxième phase du projet de modélisation prédictive du Mohawk College, où les étudiants étaient divisés en groupes expérimentaux et en un groupe témoin, afin de mieux isoler l'efficacité de la modélisation prédictive (Finnie et coll., 2017b).

Cependant, parmi ceux qui ont signalé un effet, il s'agissait presque universellement d'un effet positif; aucun effet entièrement négatif n'a été constaté, et le seul répondant qui a indiqué à la fois un effet positif et négatif a cité le fait que le modèle signalait à l'occasion des personnes qui n'en avaient pas vraiment besoin. Aucun modèle prédictif ne peut être correct dans 100 % des cas et, en général, les résultats rapportés étaient positifs.

De nombreux répondants ont indiqué un changement dans la culture et la conversation sur leur campus à la suite de l'utilisation de la modélisation prédictive. Ce changement a mené à une utilisation accrue des données et des éléments probants dans la prise de décisions, particulièrement en ce qui concerne les mesures de soutien offertes et la façon dont les étudiants sont informés des mesures de soutien disponibles. Les répondants ont suggéré que le fait d'avoir plusieurs « gains » ou améliorations dans lesquels la

modélisation prédictive a joué un rôle a contribué à la fois à l'adoption de l'information du modèle pour les interventions et à la poursuite du passage à la prise de décisions éclairées par les données.

Tirer parti des avantages de la modélisation prédictive

La majorité des répondants au sondage (15 sur 25, 60 %) ont indiqué que l'utilisation de la modélisation prédictive avait entraîné certains changements dans leur établissement. Dans certains cas, ces changements portaient sur la disponibilité ou la promotion de mesures de soutien pour les étudiants, tandis que dans d'autres cas, des changements ont été apportés à la conception des programmes d'études, aux exigences d'admission ou aux politiques de rétention. Lorsqu'il y avait un effet connu sur la persévérance et la réussite scolaires, cet effet était presque universellement positif.

Parmi les interventions élaborées par les établissements, les réponses ont montré qu'il n'y avait pas qu'une seule façon pour les établissements d'apporter des changements pour accompagner la modélisation prédictive. Dans certains cas, seuls des changements à la façon de promouvoir les mesures de soutien auprès des étudiants ont été signalés, tandis qu'au moins un répondant a indiqué qu'une refonte complète du modèle de prestation d'aide scolaire aux étudiants était en cours. Ce qui ressort clairement des réponses, c'est qu'il n'existe pas de modèle prédictif qui puisse fonctionner dans l'ensemble des établissements. Les modèles doivent être construits sur la base des informations dont ils disposent (à la fois en matière d'informations disponibles et de structure de ces informations) et selon les prédictions. Étant donné que ces éléments varient souvent à l'intérieur même des établissements et d'un établissement à l'autre, l'élaboration d'un modèle prédictif doit être ancrée dans le contexte de l'établissement et constitue un processus qui exige habituellement des ressources et une attention considérables.

Orientations futures de la recherche

Un élément qui n'a pas été abordé dans la recherche est de savoir comment l'utilisation de la modélisation prédictive affecte l'accès des étudiants, en particulier pour les groupes sous-représentés. Pour ce faire, il faudrait approfondir les résultats de la modélisation et l'adoption des interventions plus en profondeur que ce qui a été fait dans l'étude actuelle. Dans les cas où la modélisation prédictive a entraîné un changement dans l'apport de nouveaux étudiants, une analyse préalable et postérieure serait également utile.

Parmi les répondants, la plupart des interventions directes auprès des étudiants impliquent la promotion de services optionnels plutôt qu'obligatoires. La nature et le contenu exacts de la promotion des services pourraient être importants, comme l'a montré la recherche en économie comportementale sur les « nudges » (Thaler et Sunstein, 2008) qui commence à s'appliquer également à l'enseignement postsecondaire (Ross, White, Wright et Knapp, 2013).

Cette étude a révélé que le nombre d'établissements utilisant la modélisation prédictive a doublé au cours des quatre dernières années et que beaucoup d'autres envisagent sérieusement la modélisation prédictive dans un proche avenir. Une étude longitudinale de suivi sur l'impact de la modélisation prédictive nous permettrait de mieux comprendre les effets à long terme sur les établissements et les étudiants. De plus, d'autres recherches dans ce domaine pourraient aider les établissements à déterminer comment utiliser

efficacement la modélisation prédictive avec l'information qu'elles recueillent actuellement et comment l'enrichir de données supplémentaires dans l'avenir. Il est important de noter qu'il est peu probable qu'il suffise de se concentrer uniquement sur les données : pour réussir à changer un établissement, la modélisation doit être présentée et contextualisée afin qu'elle puisse stimuler le développement organisationnel (Macfadyen et Dawson, 2012).

Inversement, une grande partie des répondants ont indiqué qu'ils n'utilisaient pas la modélisation prédictive (45 répondants sur 70, soit 64 % des réponses à la question 4). Une étude future sur les raisons pour lesquelles les établissements pourraient ne pas adopter ou ne pas vouloir adopter la modélisation prédictive aiderait à comprendre les obstacles à l'adoption au sein de l'enseignement supérieur canadien.

Enfin, comme la majorité du sondage utilisé dans le cadre de la présente étude portait sur les établissements qui utilisent actuellement la modélisation prédictive, le groupe qui prévoit la mettre en œuvre sous peu n'a fourni que très peu d'information, qu'il s'agisse d'étudier ou d'envisager sérieusement d'étudier la modélisation prédictive. Un ensemble plus riche de réponses de ce groupe quant aux raisons pour lesquelles ils s'orientent vers la modélisation prédictive, aux obstacles qu'ils rencontrent et aux possibilités qu'ils entrevoient les aiderait à comprendre de façon plus globale son processus d'adoption.

L'analyse documentaire a permis de déterminer les orientations futures de la recherche, notamment l'élaboration d'une éthique de la modélisation prédictive pour la persévérance scolaire (Gašević et coll., 2016), la création d'un cadre général pour discuter de la modélisation prédictive (Chatti, Dyckhoff, Schroeder et Thüs, 2013) et d'autres recherches sur les interventions et leur incidence sur les étudiants (Jayaprakash et coll., 2014).

Limites

Il y a certaines limites à cette recherche qui doivent être reconnues. En mettant l'accent sur les établissements canadiens, cette recherche aide à combler un vide par rapport à la compréhension de la modélisation prédictive dans le contexte canadien, mais ne nous a pas permis de comparer ce que font les établissements canadiens avec la communauté internationale en général.

Aussi, il est important de noter que le présent rapport se limite au point de vue des personnes qui ont volontairement participé à l'étude et n'est donc pas représentatif des établissements canadiens (et de leur personnel). Le présent rapport se veut plutôt de nature exploratoire et descriptive afin de mieux comprendre le paysage de la modélisation prédictive au Canada.

Conclusions

La modélisation prédictive gagne en popularité comme moyen d'améliorer à la fois la planification des établissements et la persévérance scolaire. À l'aide d'une approche en deux étapes, comprenant un sondage pour recueillir un large éventail de réponses, suivi d'un questionnaire par entrevue ou par courriel pour approfondir les expériences de ceux qui ont utilisé la modélisation prédictive dans leur établissement, nous avons examiné l'utilisation de la modélisation prédictive pour la persévérance scolaire dans les établissements canadiens.

Parmi nos répondants, le nombre d'établissements utilisant la modélisation prédictive aux fins de la persévérance scolaire a plus que doublé entre 2013 et 2017, les établissements signalant des changements dans la façon dont les interventions auprès des étudiants étaient promues, offertes et évaluées. Parmi les répondants, 36 % utilisaient la modélisation prédictive et 39 % ont déclaré l'envisager sérieusement. Les méthodes de mise en œuvre de la modélisation prédictive sont très variées, allant de nombreuses techniques et sources de données différentes à des modèles pour différents groupes d'étudiants, en passant par la façon dont l'information est mise à la disposition du personnel, des chargés de cours et, dans deux cas, des étudiants eux-mêmes.

L'une des utilisations les plus novatrices de la modélisation prédictive était d'influencer les demandeurs qui recevraient une offre d'espace en résidence. De plus, un certain nombre de répondants avaient établi un lien précis entre leur utilisation de la modélisation prédictive et leurs plans stratégiques d'inscription et les besoins des collectivités locales; dans un cas particulier, le comité de gestion stratégique des inscriptions a participé activement à l'examen annuel du modèle.

Cette étude laisse des possibilités de recherches futures. Nous n'avons pas examiné comment la modélisation prédictive influe sur l'accès, ni les effets de la promotion de services de soutien obligatoires aux étudiants plutôt qu'optionnels. Étant donné que bon nombre des modèles dont il est question dans le présent rapport sont relativement nouveaux, une étude de suivi visant à déterminer les effets à plus long terme serait utile.

Références

- AACRAO (n.d.). AACRAO –American Association of Collegiate Registrars and Admissions Officers. Extrait de <http://www.aacrao.org>.
- ACPRI (n.d.). Accueil | ACPRI. Extrait de <http://cirpa-acpri.ca/fr/>.
- AIR (n.d.). Association for Institutional Research. Extrait de <http://www.airweb.org/pages/default.aspx>.
- AMOSSHE (n.d.). AMOSSHE - Accueil. Extrait de <https://www.amosshe.org.uk>.
- ANZSSA. 2015 Accueil. Extrait de <https://www.anzssa.com>.
- Arnold, K. E., et M. D. Pistilli. Course signals at Purdue, dans 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, 2012.
- ARUCC. (n.d.). Bienvenue – ARUCC – Association des registraires des universités et collèges du Canada. Extrait de <https://www.arucc.ca/fr/>.
- ASEUCC (n.d.). Association des services aux étudiants des universités et collèges du Canada. Extrait de <https://www.cacuss.ca/fr/index.html>.
- Astin, A. W. *What matters in college?: Four critical years revisited*. San Francisco : Jossey-Bass, 1993.
- Baker, R. S. et K. Yacef. The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3-17, 2009.
- Bean, J. P. et B. S. Metzner. A conceptual model of nontraditional undergraduate student attrition, *Review of Educational Research*, 55(4), 485-540, 1985.
- Campbell, J. P. *Utilizing student data within the course management system to determine undergraduate student academic success: An exploratory study* (Doctoral dissertation). Extrait de ProQuest. (3287222.), mai 2007.
- Campbell, J. P., P. B. deBlois et D. G. Oblinger. Academic analytics: A new tool for a new era. *EDUCAUSE Review*, 42(4), 40-57, extrait de <http://er.educause.edu/articles/2007/7/academic-analytics-a-new-tool-for-a-new-era>, 6 juillet 2007.
- Chatti, M. A., A. L. Dyckhoff, U. Schroeder et H. Thüs. A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5–6), 318–331. doi:10.1504/IJTEL.2012.051815, 2013.
- Cohen, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 20 (1): 37–46. Doi:10.1177/001316446002000104,:10.1177/001316446002000104, 1960.
- Dangeti, P. *Statistics for machine learning*. Birmingham, Royaume-Uni, Packt Publishing, 2017.
- Daniel, B. Big Data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, 46(5), 904–920. Doi:10.1111/bjet.12230, 2015.

- Davis, C. M., J. M. Hardin, T. Bohannon et J. Oglesby. Data mining applications in higher education. Dans K. D. Lawrence, S. Kudyba et R. K. Klimberg (dir. publ.), *Data Mining Methods and Application* (p. 123-148), 2007.
- Delen, D. 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, 2010.
- Dolence, M.G. *Strategic enrollment management: A primer for campus administrators*. Washington, DC : American Association of Collegiate Registrars and Admission Officers, 1993.
- Ellis, R. K. Field guide to learning management systems. ASTD Learning Circuits, 2009.
- Finlay, S. *Predictive analytics, data mining and big data: Myths, misconceptions and methods*. Londres : Springer, 2014.
- Finnie, R., S. Childs et A. Wismer. *Under-represented groups in postsecondary education in Ontario: evidence from the Youth in Transition Survey*. Toronto, Conseil ontarien de la qualité de l'enseignement supérieur, 2011.
- Finnie, R., T. Fricker, E. Bozkurt, W. Poirier et D. Pavlic. *Using predictive modelling to inform early alert and intrusive advising interventions and improve retention*. Toronto, Conseil ontarien de la qualité de l'enseignement supérieur, 2017. [Cité dans le document en tant que : Finnie et coll. 2017a]
- Finnie, R., T. Fricker, E. Bozkurt, W. Poirier, D. Pavlic et M. Pratt. *Academic advising: Measuring the effects of "proactive" interventions on student outcomes*. Toronto, Conseil ontarien de la qualité de l'enseignement supérieur, 2017. [Cité dans le document comme Finnie et coll., 2017b.]
- Gašević, D., S. Buckingham Shum, K. Nelson, S. Alexander, L. Lockyer, G. Kennedy et coll. *Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement*. Sydney, NSW: Australian Office of Learning & Teaching. Extrait de http://www.olt.gov.au/system/files/resources/SP13_3249_Dawson_Report_2016.pdf, 2016.
- Hoffmann, E. et M. Chamie. Standard statistical classifications : basic principles. *Statistical Journal of the United Nations Economic Commission for Europe*, 19(4), 223–241, 2002.
- Horvli, T. *Data warehouse presentation*. Consulté à l'adresse : <https://pdfs.semanticscholar.org/presentation/1747/aa686c37ae3e4171e32a96b3eb996e366906.pdf>, 2004.
- Hossler, D. et J. Bean. *The strategic management of college enrollments*. San Francisco, CA : Jossey-Bass, 1990.
- Jayaprakash, S. M., E. W. Moody, E. J. Lauría, J. R. Regan et J. D. Baron. Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics*, 1(1), 6–47, 2014.
- Jia, P. et T. Maloney. 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, 2015.
- Kahu, E. R. Framing student engagement in higher education. *Studies in Higher Education*, 38(5), 758-773, 2013.

- Kuh, G. D., T. M. Cruce, R. Shoup, J. Kinzie, et R. M. Gonyea. Unmasking the effects of student engagement on first-year college grades and persistence. *The Journal of Higher Education*, 79(5), 540–563, 2008.
- Kuzilek, J., M. Hlosta, D. Herrmannova, Z. Zdrahal, et A. Wolff. OU Analyse : Analysing at-risk students at The Open University. *Learning Analytics Review*, LAK15(1), 1–16, 2015.
- Lam, Y. L. J. Predicting dropouts of university freshmen: A logit regression analysis. *Journal of Educational Administration*, 22(1), 74–82. Doi:10.1108/eb009886, 1984.
- Macfadyen, L. et S. Dawson. Numbers alone are not enough: why e-learning analytics failed to inform an institutional strategic plan. *Educational Technology and Society*, 15(3): 149-163, 2012.
- Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, et A. Hung Byers. *Big data: The next frontier for innovation, competition, and productivity*. Extrait du site Web du McKinsey Global Institute : <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>, 2011.
- Minaei-Bidgoli, B. et W. F. Punch. Using genetic algorithms for data mining optimization in an educational web-based system. Dans E. Cantú-Paz, J. A. Foster, K. Deb, L. D. Davis, R. Roy, U.-M. O'Reilly et coll. (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, 2003.
- Morris, L. V., S.-S. Wu et C. L. Finnegan. Predicting retention in online general education courses. *The American Journal of Distance Education*, 19(1), 23–36. Doi:10.1207/s15389286ajde1901_3, 2005.
- Pascarella, E. T. et P. T. Terenzini. *How college affects students*. (K. Feldman, Ed.) (Vol. 2). San Francisco : Jossey-Bass, 2005.
- QuestionPro. (n.d.) *Sondages en ligne QuestionPro*. Extrait de <https://ca.questionpro.com/fr/>.
- Reinard, J. *Communication Research Statistics*. Thousand Oaks, CA: Sage Publications, 2006.
- Richardson, M., C. Abraham et R. Bond. Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353–387. Doi:10.1037/a0026838, 2012.
- Ross, R., S. White, J. Wright, et L. Knapp. *Using behavioral economics for postsecondary success*. Cambridge, MA : ideas42, 2013.
- Scalise, A., M. Besterfield-Sacre, L. Shuman et H. Wolfe. First term probation: Models for identifying high risk students. Dans *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, 2000.
- SocioCultural Research Consultants . *Dedoose* (Version 7.6.21)[logiciel d'application Web]. Extrait de <http://www.dedoose.com>, 2017.
- State University. *College student retention – defining student retention, a profile of successful institutions and students, theories of student departure*. Extrait de : <http://education.stateuniversity.com/pages/1863/College-Student-Retention.html>, 2018.

- Statistique Canada . La scolarité au Canada : faits saillants du Recensement de 2016. *Le Quotidien*. Extrait de <https://www150.statcan.gc.ca/n1/daily-quotidien/171129/dq171129a-fra.htm?HPA=1>, 29 novembre 2017.
- Thaler, R. H., C. S. Sunstein. *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press, 2008.
- Tinto, V. Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45(1), 89-125, 1975.
- Tinto, V. *Leaving college: Rethinking the causes and cures of student attrition*. Chicago : University of Chicago Press, 1987.
- University of Maryland University College . *Predictive analytics for student success: Developing data-driven predictive models of student success*. Extrait de <https://www.umuc.edu/documents/upload/developing-data-driven-predictive-models-of-student-success-final.pdf>, 2015.
- van Barneveld, A., K. E. Arnold et J. P. Campbell. Analytics in higher education: Establishing a common language. *EDUCAUSE Learning Initiative*, 1, 1–11. Extrait de <https://library.educause.edu/~media/files/library/2012/1/eli3026-pdf.pdf>, 2012.
- Wiggers, R. et C. Arnold. *Defining, measuring and achieving “student success” in Ontario colleges and universities* (@ Issue Paper No. 10). Toronto, Conseil ontarien de la qualité de l’enseignement supérieur. Extrait de <http://www.heqco.ca/SiteCollectionDocuments/AtIssueStudent Success ENG.pdf>, 2011.
- Zhang, Y., S. Oussena, T. Clark et H. Kim. Use data mining to improve student retention in higher education - A case study. Dans *International Conference on Enterprise Information Systems* (pp. 190-197), 2010.



Higher Education
Quality Council
of Ontario

An agency of the Government of Ontario