

Léon Harvey, Marie-Hélène Hébert et Catherine Simard

# Qualités des dossiers universitaires : analyse d'un programme de formation en enseignement

## 1. Validité et équité dans l'évaluation des étudiants

- 1 L'objectif du présent article est de démontrer que les notes qui figurent au dossier académique<sup>1</sup> d'étudiants du programme d'éducation préscolaire et d'enseignement primaire d'une université québécoise sont valides, fiables, équitables quant à l'âge et au genre des étudiants, équivalentes entre les campus et les années de diplomation, et sont, tel qu'attendu, liées aux résultats académiques antérieurs des étudiants.
- 2 L'évaluation en contexte scolaire est définie (JCSEE, 2012) comme étant un processus de collecte et d'interprétation de l'information qui est utilisée pour informer les acteurs concernés (élèves/étudiants, enseignants/professeurs, parents/tuteurs, etc.) des progrès réalisés à l'université. Une telle évaluation n'est pas en soi une mesure standard et n'est pas non plus réalisée à partir d'outils standardisés. Teintée de subjectivité, elle n'est jamais totalement neutre (Rey, 2008). Par conséquent, elle nécessite que l'évaluateur soit responsable de ses actes et qu'il engage une réflexion (Beauvais, 2012/3) ou un questionnement (Hadji, 2012) sur ses actions évaluatives. Le présent article invite à une réflexion sur le produit des évaluations : les notes académiques qui sont attribuées aux étudiants et qui figurent dans leur dossier académique.
- 3 Différents organismes d'accréditation, tant en Europe qu'aux États-Unis et au Canada, énoncent des critères de qualité vis-à-vis des évaluations. Ces organismes ont des attentes explicites face à l'évaluation des retombées de l'apprentissage. Par exemple, aux États-Unis, le CHEA (2003) considère que les institutions et les programmes sont responsables d'établir des énoncés clairs relatifs aux acquis et aux aptitudes des étudiants en fin de cursus et spécifique que des données probantes à cet effet doivent être recueillies et communiquées. En Europe, l'ENQA (2009) a publié des standards généraux destinés à l'enseignement supérieur ; le standard 1.3 stipule que les étudiants doivent être évalués en fonction de critères, de règles et de procédures qui sont appliqués de façon cohérente (voir également CAPFE, 2010 ; CSE, 2012 ; JCSEE, 2012 ; Lewis, 2011).
- 4 À ce titre, certaines politiques institutionnelles en enseignement supérieur (par exemple, Collège Marie-Victorin, 2005 ; JCSEE, 2012) explicitent les qualités attendues des notes académiques. Elles considèrent que les notes rapportées au dossier académique ne doivent pas être le résultat d'un processus purement arbitraire et aléatoire ; elles doivent plutôt être équitables, équivalentes, valides et fiables et en cohérence avec les objectifs du programme de formation. Ce faisant, elles énoncent certains axes de réflexion quant aux qualités attendues de ce produit des évaluations.
- 5 Il existe également un intérêt grandissant pour les mécanismes complexes qui régissent l'évaluation et la transparence dans l'obtention des notes afin d'assurer la qualité en enseignement (Ryan & Cousins, 2009). Des évaluations de qualité informent les étudiants des progrès réalisés lors de leurs apprentissages, participent à la régulation des apprentissages ultérieurs, et finalement, donnent des rétroactions utiles pour des fins de modifications de programmes (Lekholm & Cliffordson, 2008).
- 6 La politique sur l'évaluation des apprentissages du Ministère de l'Éducation du Québec (MEQ, 2003) définit bien la notion d'équité. Elle nécessite que soient prises en compte, dans les pratiques d'évaluation, les caractéristiques individuelles ou communes à certains groupes, afin d'éviter que, par l'évaluation, l'école ne contribue à accroître les différences existantes (MEQ, 2003). En ce sens, l'évaluation ne doit pas introduire de biais qui avantageraient ou désavantageraient certains étudiants (MEQ, 2003). À cette fin, les professeurs ne doivent évaluer que l'atteinte des objectifs prévus au programme et qui font l'objet d'activités

d'enseignement ou d'apprentissage. Ainsi, les résultats académiques sont équitables lorsqu'ils reflètent bien l'atteinte des objectifs visés par un programme de formation. Par la même logique, il est équitable que les succès académiques soient corrélés à des réalisations antérieures des étudiants, à des tests d'aptitudes standardisés concomitants ou à des réalisations professionnelles futures.

- 7 Il est, par contre, inéquitable que des différences existent si elles ne sont pas liées à l'atteinte des objectifs d'une formation, aux succès ou aux aptitudes des étudiants. Pourtant, la littérature rapporte fréquemment des différences liées au genre (Lekholm & Cliffordson, 2008 ; Luthy, 1996 ; Pfister, 1975) et à l'âge (Luthy, 1996) des étudiants. Ainsi, Jabor, Machtmes, Kungu, Buntat et Nordin (2011) soulignent que les filles performant mieux et obtiennent de meilleurs résultats académiques que les garçons à tous les niveaux du système scolaire. Il est, par conséquent, considéré inéquitable que les femmes au Canada, malgré leurs succès académiques, ne représentent que 33 % des diplômées dans le domaine des sciences, technologies, informatique, mathématiques et génie (Hango, 2013). Par ailleurs, l'âge peut être une indication qu'un étudiant a éprouvé des difficultés par le passé ; cependant, les personnes plus âgées adoptent des comportements plus matures que les plus jeunes, ce qui favorise l'obtention de meilleurs résultats académiques. Ces différences attribuables au genre et à l'âge dans la réussite scolaire universitaire ne traduisent-elles pas, pour reprendre la taxonomie de Hadji (2012), soit des iniquités dans l'accès aux ressources, dans le traitement des individus, dans l'évaluation du rendement, et au final dans la qualité du développement individuel ?
- 8 Quant à l'équivalence, elle exige que les étudiants d'un cours donné soient évalués, quelle que soit la ressource professorale, à partir d'objectifs et de critères comparables et selon une pondération semblable (Collège Marie-Victorin, 2005). L'équivalence nécessite que soient étudiées, par exemple, les évaluations différenciées qui se produisent lorsqu'un même programme est offert dans deux campus différents d'une même université, à partir de ressources différentes. Les mêmes cours du programme sont alors offerts, mais des évaluations différentes sont réalisées à l'intérieur de chacun d'eux.
- 9 Les étudiants d'un campus sont-ils alors avantagés par rapport à ceux qui étudient dans l'autre campus ? Cette situation est peu décrite. Il existe pourtant des différences documentées dans les résultats académiques entre les écoles qui offrent des programmes semblables (Ma, Ma & Bradley, 2008), différences qui sont liées soit au contexte ou au climat des institutions. Le contexte de l'école est alors défini comme étant les ressources physiques de l'école, les caractéristiques de la population étudiante (milieu socio-économique, composition ethnique) ainsi que les particularités du corps enseignant (niveau d'éducation, expérience, etc.). Le climat, quant à lui, réfère à des facteurs qui influencent les environnements d'apprentissage (politiques administratives, organisation pédagogique, valeurs, etc.).
- 10 Les travaux de Rojstaczer et Healy (2012) suggèrent qu'il existe, aux États-Unis, des différences entre les universités publiques et privées, situées au nord ou au sud, ou qui offrent des programmes techniques ou non techniques. Ainsi, à aptitudes égales, les étudiants des universités privées, du nord, inscrits dans des programmes non techniques obtiennent de meilleurs résultats que les étudiants des universités publiques, du sud, inscrits dans des programmes techniques.
- 11 L'équivalence nécessite également que soient étudiées les différences qui existent entre les cohortes qui, année après année, se voient offrir le même programme, mais à partir de ressources différentes, ce qui peut entraîner des différences dans les résultats académiques. Ici encore, les écrits rapportent des différences entre les années ainsi qu'un phénomène progressif d'inflation des notes qui opère depuis environ une trentaine d'années (Jewell & McPherson, 2012 ; Rojstaczer & Healy, 2012 ; Smith, 1992). De sorte que, pour un même secteur d'études, les diplômés récents reçoivent des notes supérieures aux diplômés antérieurs. Au fil des années, la moyenne cumulative des notes est passée de la note littérale C à B (Caruth & Caruth, 2013). Ce phénomène d'inflation des notes soulève un problème d'équivalence entre les générations ; il est une indication que des notes académiques élevées sont plus faciles à obtenir que par le passé. Cependant, il n'y a pas en soi d'effet de l'inflation des notes sur la fiabilité de la moyenne cumulative (Millman, Slovacek, Kulick & Mitchell, 1983). De

plus, un changement dans les résultats moyens ne signifie pas qu'une formation n'atteint pas ses objectifs en termes d'évaluation. Selon Pattison, Grodsky et Muller (2013), il faut plutôt s'interroger sur la valeur informative des diplômes. La valeur informative est la puissance de la moyenne cumulative à envoyer un signal adéquat (signaling power of grades) quant à la qualité des étudiants qui sont diplômés. Une augmentation de la moyenne ne change pas en soi cette valeur informative. Il faut plutôt vérifier les changements dans la variance des notes ; changements qui informent que la distance qui sépare les excellents étudiants de ceux qui sont soit très bons, moyens ou médiocres a changé avec le temps.

12 Ce qui est encore plus déterminant selon Pattison Grodsky et Muller (2013), ce sont des changements dans les covariations qui existent entre les notes académiques obtenues avec les antécédents des étudiants en termes de réussite et d'efforts ainsi que les conséquences des évaluations en termes de niveaux d'études atteints et de conditions de travail (prestige occupationnel, salaire, etc.). En ce sens, Pattison, Grodsky et Muller (2013) n'ont trouvé aucune preuve qu'il y a eu une perte dans la valeur informative de la moyenne cumulative (GPA) entre 1972 et 1992 dans les universités américaines.

13 Ceci amène à considérer deux autres qualités au moment de la lecture des dossiers académiques, soit la validité et la fiabilité. La validité est un jugement global qui est porté sur l'ensemble des preuves accumulées à propos d'un dossier académique et qui détermine si les interprétations réalisées à partir de celui-ci sont conformes à l'intention visée (AERA, 1999). Lorsqu'il s'agit de porter un jugement sur l'ensemble des notes accumulées dans le cadre d'un curriculum, la notion de validité de construit curriculaire est introduite (Harvey, 2012). Elle permet de distinguer, ce qui, dans une note, est lié à un facteur commun, de ce qui est attribuable à une erreur aléatoire ou systématique. La détection d'erreurs systématiques est généralement une indication de la présence de biais et d'iniquités. Quant à la fiabilité, elle renvoie à la qualité de l'évaluation. Une évaluation est fiable lorsqu'elle génère des erreurs aléatoires ou systématiques faibles par rapport à ce qui est noté.

14 En somme, la problématique dans le cadre d'une évaluation est de savoir le rapport entre ce qui relève de facteurs structurés et en lien avec les objectifs du programme et ce qui relève d'erreurs aléatoires ou systématiques. Au regard de cette problématique, nombre d'études ont en commun de développer des preuves quant à la validité de construit des notes académiques obtenues dans le cadre des programmes d'études étudiés (Goova, Hollett, Tesfay, Gala, Puzifferri, Kehdy & Scott, 2008 ; Mason & Dragovich, 2010 ; Rexwinkel, Haenen & Pilot, 2013). D'autres études ont considéré la validité de critère et ont montré une relation entre la moyenne cumulative, les résultats antérieurs, des tests d'aptitudes standardisés (Camara & Echternacht, 2000 ; DiPerna, 2004 ; Geiser & Santelices, 2007 ; Kobrin, Patterson, Shaw, Mattern & Barbuti, 2008 ; Komarraju, Ramsey & Rinella, 2013) ou des tests de connaissances spécifiques aux disciplines considérées (Bacon & Bean, 2006 ; Luthy, 1996 ; Smith, 1992). Aux cycles supérieurs, Kuncel, Credé et Thomas (2007) ont réalisé des métaanalyses des études en gestion des affaires, tandis que Kuncel, Wee, Serafin et Hezlett (2010) ont considéré les études de maîtrise et de doctorat, toutes disciplines confondues. Saupe et Eimers (2012) rapportent, quant à eux, que la moyenne cumulative à l'université (Grade point average, GPA) possède une bonne fiabilité.

15 Notons que, malgré la diversité des études considérées, peu d'entre elles se sont intéressées à la validité des notes académiques dans le contexte de la formation des maîtres. Or, il existe, dans le monde, plusieurs centaines de programmes de formation des maîtres dont certains sont soumis à un processus d'accréditation afin d'en assurer la qualité. Au Québec, le comité d'accréditation (CAPFE) exige notamment des preuves qui attestent que les compétences professionnelles attendues chez les étudiants au terme de ces programmes sont bel et bien évaluées. En ce sens, le comité d'agrément inspecte les plans de cours, les grilles de notation utilisées, discute avec les étudiants et les ressources professorales et considère toutes les autres preuves que les universités souhaitent produire. Il peut alors s'agir de bilans d'apprentissage, de portfolios ou d'autres productions réalisées par les étudiants. De plus, les universités mettent en place des approches-programmes qui assurent une plus grande cohérence dans les activités

des programmes, et ce, incluant les activités d'évaluation (Prégent, Bernard & Kozanitis, 2009).

16 À ces efforts, nous considérons qu'il s'avère informatif, dans le cadre de ces formations, de démontrer qu'une note n'est pas « faussée par une série – ou plutôt une convergence confuse - d'éléments fortuits et subjectifs » (Pfister, 1975, p. 11 citant Agazzi, 1967). À ce titre, une étude (Harvey, 2012) montre que les évaluations réalisées lors des cours d'un programme de formation en enseignement secondaire d'une université québécoise se structurent, tel qu'attendu, en quatre facteurs qui ont, dans l'ensemble, une fiabilité acceptable (alpha de Cronbach = 0,69). Cette étude n'a cependant pas envisagé les questions relatives à l'équité entre les étudiants (genre et âge) et l'équivalence entre les campus d'une même université ainsi que les fluctuations possibles entre les années de diplomation.

17 Dans le présent article, les notes qui figurent au dossier académique des finissants du programme d'éducation préscolaire et d'enseignement primaire d'une université québécoise sont revisitées afin de vérifier si elles s'avèrent valides, fiables, équitables quant à l'âge et au genre des étudiants, équivalentes entre les campus d'une même université et les années de diplomation, et sont, tel qu'attendu, liées aux résultats académiques antérieurs des étudiants.

## 2. Une analyse des dossiers académiques

18 L'échantillon de notre enquête se compose de 873 finissants du programme d'éducation préscolaire et d'enseignement primaire ayant gradué entre 2006 et 2012 au campus de Rimouski (n = 430) ou de Lévis (n = 443) de l'Université du Québec à Rimouski. Ils représentent 14 groupes d'étudiants (sept groupes dans chacun des campus). Le nombre de diplômés en fonction des années 2006 à 2012 est respectivement de 163, 141, 127, 131, 111, 125 et 75. Une forte majorité (92,55 %) de cet échantillon se compose de femmes (808 femmes pour 65 hommes). L'âge moyen au dernier trimestre est de 24,41 ans avec un écart-type de 3,18 et une étendue se situant entre 21,00 et 45,00 ans.<sup>2</sup> La cote de rendement au collège (CRC) des étudiants est utilisée comme indicateur des succès antérieurs. La CRC moyenne des candidats en provenance du niveau collégial (n = 835) est de 26,16 avec un écart-type de 3,21 et une étendue entre 17,58 et 35,02. Aucune différence significative (t = 0,61, p = 0,54) n'existe entre les CRC des étudiants des campus de Lévis (moyenne = 26,23, n = 424) et de Rimouski (moyenne = 26,09, n = 409). Aucune CRC n'est disponible pour les candidats admis sur une base adulte (n = 22) ou ayant l'équivalent d'un diplôme au collège (n = 16).

### 2.1. La description du curriculum

19 La scolarité du programme d'éducation préscolaire et d'enseignement primaire de notre étude s'échelonne sur quatre années. Le programme a été implanté en 2002. Il a subi de légères modifications en 2003, 2006 et 2012. La liste des cours retenus pour la présente étude apparaît dans le tableau 1. Ces cours forment la majeure partie du corpus obligatoire du programme qui se compose de 120 unités et prépare les futurs enseignants à enseigner au préscolaire et au primaire. Tous les cours sont des activités de trois unités (45 heures de cours et 145 heures de travail personnel). Les quatre cours de formation pratique où une notation succès-échec est utilisée ne sont pas considérés dans l'analyse qui nécessite des variables dépendantes continues.

### 2.2. Les compétences évaluées

20 Les objectifs visés par le programme de formation se présentent sous la forme de 12 compétences. Il existe plusieurs définitions de ce concept dans la francophonie. Ayotte-Beaudet (2013) en recense environ une vingtaine. L'aspect central est la mobilisation par un individu d'un ensemble de ressources internes et externes afin de réaliser des tâches authentiques issues de familles de situations. Par définition, une compétence est non directement observable. Cependant, sa structure peut être inférée à partir d'indicateurs, qui eux, sont observables. Dans la présente étude, les notes académiques permettront d'inférer la structuration des compétences dans le programme étudié. En théorie, ces compétences se structurent en quatre groupes distincts qui sont décrits dans le référentiel du Ministère de l'Éducation du Québec (MEQ, 2001). Les compétences sont numérotées C1 à C12 pour

faciliter leur identification. Le groupe *Fondements* concerne la culture de l'éducation (C1) et la communication orale et écrite (C2). L'*Acte d'enseigner* vise la conception (C3) et le pilotage des situations pédagogiques (C4), l'évaluation des compétences (C5), et la gestion du groupe-classe (C6). Les *Aspects sociaux* mobilisent l'adaptation de l'enseignement (C7), les technologies de l'information et de la communication (C8), la coopération avec l'équipe-école (C9) et l'équipe pédagogique (C10). Finalement, le groupe *Identitaire* inclut le développement (C11) et l'éthique (C12) professionnels. Le programme de formation précise comment les cours concourent à l'évaluation de ces compétences. Les ressources professorales qui offrent les cours sont responsables de s'assurer que ces objectifs de formation sont atteints<sup>3</sup>.

### 2.3. Le corpus de données

- 21 La base de données a été préparée par le Service de la recherche institutionnelle de l'Université du Québec. Elle contient deux types de variables dépendantes, soit 1) le résultat obtenu dans chacun des cours du programme par chacun des étudiants diplômés entre les années 2006 et 2012 et 2) la moyenne cumulative de l'étudiant calculée sur l'ensemble des cours suivis dans le cadre du programme. En tout, les résultats obtenus dans 28 cours (voir tableau 1) forment les variables qui sont soumises à une analyse factorielle et à l'analyse de la fiabilité. La moyenne cumulative est la variable dépendante dans le cadre d'analyses multiniveaux ; les variables indépendantes alors considérées sont l'année de diplomation, le genre, l'âge à la fin des études, la cote de rendement au collège et le campus d'études (Rimouski ou Lévis) où le programme a été suivi.
- 22 La base de données est imparfaite. Elle compte en moyenne 4,15 % de données manquantes pour chacun des cours, avec une étendue se situant entre 1 % et 16 %. Les données manquantes proviennent soit de l'obtention d'une équivalence pour un cours suivi dans une autre université ou du fait que la note a été attribuée par la ressource professorale après la dernière mise à jour des résultats pour un cours donné.
- 23 L'impact de ce 4 % de données manquantes est variable en fonction des analyses à effectuer dans le cadre de la présente étude. À titre explicatif, pour le calcul des statistiques descriptives, les sujets pour lesquels une donnée manquante existe sur la variable analysée sont simplement ignorés du calcul. Pour le calcul des coefficients de corrélation entre deux variables, ainsi que lors des analyses de régression, les calculs sont effectués en éliminant les sujets pour lesquels une donnée manquante existe sur l'une ou l'autre des deux variables corrélées (*pairwise deletion*). Pour l'analyse factorielle, la procédure par défaut élimine les étudiants si une donnée manquante existe sur l'une ou l'autre des 28 variables considérées (*listwise deletion*). Dans ce cas, 408 des 873 sujets seraient omis en raison de données manquantes, ce qui réduit considérablement la puissance de l'analyse.
- 24 Par conséquent, pour cette analyse, les données manquantes ont été remplacées en utilisant la stratégie d'imputations multiples, considérée la meilleure stratégie de substitution. Pour ce faire, la procédure MI du progiciel SAS 9.2 a été utilisée. Celle-ci crée des bases de données multiples dans lesquelles les données manquantes sont remplacées par des valeurs obtenues à partir de simulations de Monte-Carlo. Aux fins des présentes analyses, cinq bases de données alternatives furent créées par imputations multiples. Ces cinq bases de données, numérotées de 1 à 5 pour les distinguer, furent alors analysées de la même façon que la base initiale de données. Les résultats des six bases de données sont ainsi rapportés.

### 2.4. L'analyse des données

- 25 La démarche s'inspire d'études antérieures (Harvey, 2012 ; Harvey, Hébert, & Simard, 2013 ; Rexwinkel, Haenen & Pilot, 2013). Les deux premières étapes consistent à faire une description des notes académiques. Sont décrits les moyennes et écarts-types pour chacun des cours (étape 1) ainsi que le traitement des données manquantes (étape 2). À l'étape 1, l'idée de base est de déterminer si la distribution des notes obtenues dans les cours s'avère discriminante. Le critère de Rexwinkel, Haenen et Pilot (2013) est que les écarts-types ne soient ni trop faibles ni trop élevés. Lors de l'inspection des résultats académiques, sur une échelle à 10 échelons (de 1, vraiment faible à 10, excellent), Rexwinkel, Haenen et Pilot (2013) considèrent que des cours avec un écart-type d'environ 0,70 sont acceptables. À l'opposé, des

écarts-types inférieurs à 0,34 ou supérieurs à 1,40 sont respectivement considérés trop faibles et trop accentués<sup>4</sup>. À l'étape 2, des comparaisons entre les moyennes et les écarts-types de chacune des bases de données sont réalisées afin de démontrer l'équivalence des bases de données simulées avec la base originale.

26 Les étapes 3 et 4 s'insèrent dans une démarche associée à la validation de construit curriculaire. La logique est développée par Harvey (2012). Au départ, les construits notés dans les cours sont non observables et considérés latents. Néanmoins, la note obtenue dans chaque cours par les étudiants est un indicateur de la présence plus ou moins forte de ces construits latents<sup>5</sup>.

27 Par conséquent, la matrice de corrélation entre les cours du programme contient des informations précieuses sur les construits communs qui sont notés dans le cadre des cours. Cette matrice de corrélation sera générée à l'étape 3 pour la base de données initiale afin de déterminer s'il existe des relations entre les notes obtenues dans les cours. Des corrélations positives indiquent que des construits communs sont notés ; des corrélations nulles sont acceptables dans l'éventualité où il est prévu, dans le curriculum, que des cours n'évaluent pas de construits communs ; finalement, la présence de corrélations négatives est problématique, car elles indiquent que ce qui est valorisé et évalué dans un cours ne l'est pas nécessairement dans l'autre. À cet effet, toute corrélation négative devrait faire l'objet d'explications de la part des responsables du programme (Harvey, Hébert & Simard, 2013).

28 À l'étape 4, la structuration factorielle des notes obtenues dans les cours est extraite à partir d'une analyse factorielle exploratoire afin de révéler la mesure de construits latents communs. Une structure factorielle forte, où les valeurs propres (*eigenvalues*) supérieures à 1 expliquent plus de 50 % de la variance, est recherchée (Rexwinkel, Haenen & Pilot, 2013). Les cinq bases de données générées par imputations multiples seront utilisées, ainsi que la base initiale.

29 L'étape 5 est une étude de la fiabilité à partir du coefficient alpha de Cronbach où les notes obtenues dans l'ensemble des cours doivent démontrer une fiabilité supérieure à 0,60 (Rexwinkel, Haenen & Pilot, 2013). La base de données originale est utilisée ici.

30 Les étapes 6, 7 et 8 sont réalisées à partir de modèles de régression multiniveaux. Ces modèles permettent de vérifier l'effet global des covariables sur la moyenne cumulative. La base de données originale sera utilisée à cette fin. L'étape 6 permet de vérifier la validité de critère, à savoir si les notes obtenues sont liées aux résultats académiques antérieurs, c'est-à-dire à la cote de rendement au collège. Il n'y a pas de critère explicite sur la magnitude de cette relation. Basée sur les écrits recensés, nous considérons néanmoins que les résultats antérieurs devraient permettre d'expliquer au moins 30 % de la variance des résultats académiques.

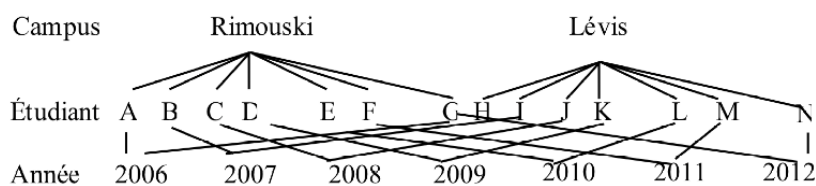
31 À l'étape 7, l'équité entre les individus sera testée en fonction du genre et de l'âge des étudiants. Finalement, à l'étape 8, l'équivalence quant au campus d'études et à l'année de diplomation sera vérifiée. Afin d'assurer qu'il y a équité et équivalence, l'absence d'effet de ces variables sur la moyenne cumulative est recherchée comme critère de qualité. Considérant que les étapes 6, 7 et 8 nécessitent l'utilisation de modèles complexes, la section suivante se concentre sur l'explicitation de la démarche d'élaboration des modèles de régression multiniveaux utilisés dans l'analyse des données.

32 La validité de critère sera examinée à partir de la capacité des résultats antérieurs à prédire le succès universitaire. De plus, les différentes covariables sont insérées afin d'identifier des sources d'iniquités possibles. Les variables considérées au niveau individuel sont la cote de rendement au collège (CRC), le genre et l'âge des étudiants. Le genre est codifié 0-1 respectivement pour les hommes et les femmes. L'âge et la CRC sont des variables continues et ne sont pas recodées. Les covariables qui distinguent les cohortes sont l'année de diplomation (2006 à 2012) et le campus d'études (Rimouski et Lévis). Ces variables sont considérées comme des variables de classification et ne nécessitent pas d'être recodées.

33 Les régressions à multiples niveaux sont réalisées à partir de la procédure Proc Mixed de SAS 9.2. Elles concernent les 833 sujets qui possèdent des données complètes sur chacune des variables. Par conséquent, 40 sujets, pour qui aucune CRC ne figure au dossier, sont exclus. Il s'agit des 38 candidats admis sur la base adulte ou ayant un diplôme équivalent et de deux étudiants en provenance du collège pour qui la CRC est manquante.

34 Les données présentent une structure hiérarchique croisée (Beretvas, 2008), qui est illustrée à la figure 1. Le caractère croisé de la structure provient du fait que les étudiants ayant diplômé une année donnée peuvent provenir d'un campus ou l'autre. Beretvas (2008) présente en détail les modèles de régression à multiples niveaux s'appliquant à l'analyse d'une telle structure. Sa notation sera utilisée dans les équations suivantes.

**Figure 1. Structure hiérarchique croisée.**



35 Afin d'explorer adéquatement cette structure, trois modèles, dont la complexité est croissante, seront développés. La comparaison entre ces trois modèles permettra d'apprécier la pertinence liée à l'ajout de variables supplémentaires en comparant les indices d'adéquation (Déviance et BIC, voir McCoach & Black (2008) pour une définition dans le cadre d'un modèle à multiples niveaux).

36 Au niveau 1, le premier modèle dit inconditionnel (parfois appelé *modèle nul*) se présente comme suit :

|  |  |       |
|--|--|-------|
|  | $Y_{i(j_1,j_2)} = \Pi_{0(j_1,j_2)} + e_{i(j_1,j_2)}$ | (1.0) |
|--|--|-------|

37 où  $Y_{i(j_1,j_2)}$  représente la moyenne cumulative de l'étudiant  $i$  provenant du campus  $j_1$  et ayant diplômé l'année  $j_2$ . La constante  $\Pi_{0(j_1,j_2)}$  représente la prédiction moyenne pour les étudiants d'un campus et d'une année donnée. L'erreur  $e_{i(j_1,j_2)}$  est la déviation propre à chaque étudiant de la moyenne de son campus et de son année de diplomation. L'erreur est présumée normalement distribuée de moyenne égale à 0 et de variance  $\sigma^2$ . Un test de normalité de distribution par le Khi-carré sera effectué sur les erreurs pour vérifier cet aspect du modèle.

38 Au niveau 2, les effets liés aux groupes d'appartenance de l'étudiant sont utilisés pour prédire la valeur de la constante  $\Pi_{0(j_1,j_2)}$  :

|  |   |       |
|--|---|-------|
|  | $\Pi_{0(j_1,j_2)} = \theta_{000} + b_{0j_10} + c_{00j_2}$ | (2.0) |
|--|---|-------|

39 où  $\theta_{000}$  est la moyenne des moyennes cumulatives,  $b_{0j_10}$  est l'effet associé au campus d'études  $j_1$ , et  $c_{00j_2}$ , celui associé à l'année d'études  $j_2$ . Les effets *Campus* ( $b_{0j_10}$ ) et *Année* ( $c_{00j_2}$ ) sont, en principe, normalement distribués de moyenne 0. Les variances respectives sont  $\#_{b00}$  et  $\#_{c00}$ .

40 L'introduction des variables indépendantes se fera progressivement au niveau 1 par l'ajout de deux modèles  $M_1$  et  $M_2$ . Dans le modèle  $M_1$ , la CRC est introduite dans l'équation (1.0) :

|  |  |       |
|--|--|-------|
|  | $Y_{i(j_1,j_2)} = \Pi_{0(j_1,j_2)} + \Pi_{1(j_1,j_2)} X_{i(j_1,j_2)} + e_{i(j_1,j_2)}$ | (3.0) |
|--|--|-------|

41 où  $X_{i(j_1,j_2)}$  est la CRC de l'étudiant  $i$  du campus  $j_1$  et de l'année  $j_2$ .  $\Pi_{1(j_1,j_2)}$  est le pondérateur associé.

42 Le modèle  $M_2$  ajoute l'effet des variables liées au genre et à l'âge des diplômés et se présente comme suit :

|  |  |       |
|--|--|-------|
|  | $Y_{i(j_1,j_2)} = \Pi_{0(j_1,j_2)} + \Pi_{1(j_1,j_2)} X_{i(j_1,j_2)} + \Pi_{2(j_1,j_2)} Z_{i(j_1,j_2)} + \Pi_{3(j_1,j_2)} W_{i(j_1,j_2)} + e_{i(j_1,j_2)}$ | (4.0) |
|--|--|-------|

43 où  $Z_{i(j_1,j_2)}$  est le genre et  $W_{i(j_1,j_2)}$ , l'âge du diplômé. Dans le présent modèle, les pondérateurs associés aux variables  $\Pi_{1(j_1,j_2)}$ ,  $\Pi_{2(j_1,j_2)}$  et  $\Pi_{3(j_1,j_2)}$  sont considérés comme indépendants du campus d'études et de l'année de diplomation. Par conséquent, ils sont considérés comme des effets fixes. Les interactions complexes entre ces variables n'ont pas été introduites dans ce modèle. En fait, la modélisation des interactions entre la CRC, l'âge et le genre de l'étudiant, ainsi que l'inclusion d'effets aléatoires ont provoqué une dégradation de l'indice d'adéquation ; par conséquent, ces modèles plus complexes, qui s'avèrent ainsi inutiles pour la mise en valeur

d'effets supplémentaires, ne seront pas présentés ici. La notation peut donc être simplifiée comme suit :

|  |  |       |
|--|--|-------|
|  | $Y_{i(j1,j2)} = \Pi_0(j1,j2) + \Pi_1 X_i + \Pi_2 Z_i + \Pi_3 W_i + e_{i(j1,j2)}$ | (5.0) |
|--|--|-------|

### 3. Résultats

#### 3.1. Étape 1 : l'inspection de la distribution

44 La moyenne cumulative de tous les cours de l'ensemble des étudiants est de 3,39 (B+)<sup>6</sup> sur 4,3 avec un écart-type de 0,36. L'étendue se situe entre 2,38 et 4,27. Le tableau 1 montre que la moyenne des cours se situe entre 2,61 (C+) et 3,82 (A-). Les écarts-types se situent entre 0,48 et 0,87 et suggèrent qu'il y a, dans l'ensemble des cours, une grande dispersion des résultats<sup>7</sup>.

#### 3.2. Étape 2 : le traitement des données manquantes

45 Les cinq bases de données générées par imputations multiples présentent des statistiques descriptives similaires au niveau de la moyenne et de l'écart-type pour chacun des cours. Les bases de données pourront donc être utilisées pour l'analyse factorielle (étape 4).

**Tableau 1. Statistiques descriptives pour chacun des cours**

| Sigle    | Intitulé du cours  | n   | moy. | é.-t. |
|----------|--|-----|------|-------|
| ATE40002 | Approfondissement d'un thème en éducation  | 768 | 3,72 | 0,67  |
| DIA13002 | Français au préscolaire et au premier cycle du primaire  | 845 | 3,00 | 0,81  |
| DIA15003 | Mathématique au préscolaire et au premier cycle du primaire  | 852 | 2,70 | 0,86  |
| DIA23013 | Enseignement et apprentissage de la grammaire au primaire  | 848 | 3,17 | 0,74  |
| DIA24002 | Géographie, histoire et éducation à la citoyenneté au primaire                                       | 841 | 3,53 | 0,51  |
| DIA25003 | Mathématique au deuxième cycle du primaire   | 856 | 2,61 | 0,78  |
| DIA28202 | Science et technologie au primaire   | 859 | 3,21 | 0,69  |
| DIA33003 | Français aux deuxième et troisième cycles du primaire  | 854 | 3,36 | 0,76  |
| DIA35003 | Mathématique au troisième cycle du primaire  | 855 | 2,70 | 0,77  |
| DIA43003 | Français et univers social   | 855 | 3,58 | 0,64  |
| DIA45002 | Mathématique, science et technologie   | 857 | 3,29 | 0,69  |
| PPE10002 | Projet personnel de formation: élaboration   | 864 | 3,63 | 0,59  |
| PPE11002 | Éducation et développement de l'enfant de zéro à six ans   | 820 | 2,83 | 0,82  |
| PPE11102 | Éducation et développement de l'enfant de six à douze ans  | 738 | 2,99 | 0,87  |
| PPE18002 | Théories de l'éducation et pratiques pédagogiques contemporaines                                     | 830 | 3,12 | 0,86  |
| PPE22002 | Éducation préscolaire  | 853 | 3,34 | 0,70  |
| PPE22102 | Gestion éducative de la classe au préscolaire  | 857 | 3,54 | 0,61  |
| PPE28102 | Apprentissage et enseignement  | 729 | 3,22 | 0,75  |
| PPE32203 | Activités d'éveil au préscolaire   | 867 | 3,82 | 0,53  |
| PPE36002 | Gestion éducative de la classe au primaire   | 853 | 3,58 | 0,60  |
| PPE37002 | Évaluation formative, plan de réussite et pédagogie différenciée                                     | 857 | 3,40 | 0,63  |
| PPE40012 | Projet personnel de formation: synthèse  | 857 | 3,70 | 0,61  |
| SCE10002 | Organisation de l'éducation au Québec  | 810 | 3,61 | 0,52  |
| SCE10102 | Utilisation pédagogique des technologies de l'information et de la communication en éducation (TICE) | 844 | 3,59 | 0,68  |
| SCE20002 | Connaissance et intégration des élèves handicapés et en difficulté                                   | 819 | 3,17 | 0,72  |
| SCE20202 | Prévention, adaptation et réussite scolaire  | 837 | 3,73 | 0,48  |
| SCE20303 | Interventions auprès des élèves en trouble du comportement au primaire                               | 848 | 3,63 | 0,59  |
| SCE30102 | Évaluation des compétences   | 855 | 3,37 | 0,77  |

#### 3.3. Étape 3 : la corrélation entre les cours

46 La matrice de corrélation compte 378 coefficients<sup>8</sup>. Par exemple, les cours de didactique du français DIA13002 et DIA23013 sont corrélés à 0,43 ; ceci s'explique par le fait qu'ils mobilisent des construits similaires (connaissances et compétences). À l'opposé, le même



cours de didactique du français DIA13002 n'est que faiblement corrélé ( $r = 0,12$ ) avec le cours de didactique des sciences et technologies DIA28202 ; l'explication est que ces cours n'évaluent pas la mobilisation des mêmes construits. Par contre, ce dernier est corrélé ( $r = 0,41$ ) avec le cours de didactique des mathématiques DIA25003 ; une fois encore, l'interprétation est qu'ils évaluent des construits communs.

47 La corrélation moyenne entre les notes obtenues dans les cours est de 0,30 avec un écart-type de 0,11. L'étendue se situe entre -0,04 et 0,58. Seulement neuf coefficients de corrélation sur 378 ne sont pas significativement différents de zéro. Il y a donc 97,6 % des coefficients qui sont significatifs et positifs (aucun coefficient négatif n'est significatif). À la lumière de l'inspection de la matrice de corrélation, il est possible d'affirmer que l'ensemble des 28 cours mobilisent des construits communs. L'analyse factorielle permettra d'explorer plus en profondeur la structuration de ces construits communs. Néanmoins, à ce stade-ci, le critère énoncé précédemment est atteint ; il existe des corrélations positives entre les cours et aucune corrélation négative significative n'est présente.

### 3.4. Étape 4 : l'étude du construit

48 L'analyse factorielle exploratoire appliquée sur la base initiale de données suggère la présence de cinq construits qui expliquent 55,6 % de la variance totale. Les valeurs propres associées à ces construits sont respectivement de 9,02, 2,56, 1,65, 1,28 et 1,06 (voir tableau 2). Le premier construit explique à lui seul 32,2 % de la variance totale. Tous les cours saturent dans ce construit, et ce, avec des pondérations qui se situent entre 0,38 et 0,70.

**Tableau 2. Solution factorielle avant rotation**

| Valeurs      | Initiale     | 1            | 2            | 3            | 4            | 5            |
|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| F1           | 9,02         | 9,16         | 9,22         | 9,16         | 9,28         | 9,20         |
| F2           | 2,56         | 2,61         | 2,60         | 2,64         | 2,61         | 2,64         |
| F3           | 1,65         | 1,46         | 1,46         | 1,46         | 1,46         | 1,48         |
| F4           | 1,28         | 1,19         | 1,20         | 1,20         | 1,18         | 1,21         |
| F5           | 1,06         | 1,01         | ----         | 1,01         | ----         | ----         |
| <b>Total</b> | <b>15,57</b> | <b>15,43</b> | <b>14,48</b> | <b>15,47</b> | <b>14,53</b> | <b>14,53</b> |
| %            | 55,6 %       | 55,1 %       | 51,7 %       | 55,3 %       | 51,9 %       | 51,9 %       |

49 Les solutions générées par imputations multiples donnent des solutions qui se composent soit de quatre ou de cinq facteurs. Trois bases de données (initiale, 1 et 3) sur les six considérées génèrent des solutions à cinq facteurs qui expliquent plus de 55 % de la variance totale. Les trois autres bases de données (2, 4 et 5) génèrent des solutions à quatre facteurs qui expliquent alors environ 52 % de la variance totale. Il y a donc un léger impact de la stratégie d'imputations multiples sur la sélection d'une solution à quatre ou cinq facteurs. Néanmoins, quelle que soit la base de données sélectionnée, toutes les solutions atteignent le critère de Rexwinkel, Haenen et Pilot (2013) qui consiste à obtenir une solution factorielle forte qui explique plus de 50 % de la variance totale.

50 Pour la suite des analyses et afin d'en faciliter l'interprétation, une rotation *equamax* a été effectuée sur les quatre premiers facteurs. Le choix de ne retenir que quatre facteurs est justifié par une perspective théorique, car le programme de formation prévoit la mesure de quatre groupes distincts de compétences (MEQ, 2001). De plus, si les analyses réalisées jusqu'à présent permettent de confirmer la présence de quatre facteurs, la mesure d'un cinquième facteur ne peut être confirmée par toutes les bases de données. Par ailleurs, une rotation *equamax* répartit la variance également entre les facteurs. Ici encore, ce choix est motivé par une considération théorique, car en principe, le curriculum doit développer dans les mêmes proportions les quatre groupes de compétences.

- 51 Le tableau 3 montre que les solutions factorielles avec rotation expliquent sensiblement la même proportion de variance pour chacune des bases de données. Le tableau 4 présente la composition de chacun des facteurs. Les saturations moyennes des bases de données 1 à 5 sont présentées. Les saturations inférieures à 0,30 ne sont pas représentées (seule celle du cours DIA15003 égale à 0,29 a été retenue, car ce cours s'insère logiquement dans le premier facteur).

**Tableau 3. Solution factorielle après rotation *equamax***

| Valeurs      | Initiale     | 1            | 2            | 3            | 4            | 5            |
|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| F1           | 4,37         | 4,20         | 4,20         | 4,25         | 4,22         | 4,21         |
| F2           | 3,61         | 3,68         | 3,72         | 3,64         | 3,69         | 3,71         |
| F3           | 3,30         | 3,50         | 3,56         | 3,53         | 3,49         | 3,56         |
| F4           | 3,24         | 3,05         | 3,00         | 3,03         | 3,12         | 3,04         |
| <b>Total</b> | <b>14,52</b> | <b>14,43</b> | <b>14,48</b> | <b>14,45</b> | <b>14,52</b> | <b>14,52</b> |
| %            | 51,9 %       | 51,5 %       | 51,7 %       | 51,6 %       | 51,9 %       | 51,9 %       |

- 52 Le premier facteur apparaît comme un facteur de fondements de l'éducation et de didactique des mathématiques, sciences et technologies. Ainsi, il regroupe les quatre cours de didactique des mathématiques DIA15003, DIA25003, DIA35003 et DIA45002 ainsi que le cours de didactique des sciences et technologies (DIA28202). S'ajoutent des cours en fondements de l'éducation portant sur le développement de l'enfant (PPE11002 et PPE11102), l'apprentissage, l'enseignement et les théories éducatives (PPE18002 et PPE28102), l'évaluation formative (PPE37002) ainsi qu'un cours sur la connaissance des élèves handicapés (SCE20002). S'insèrent finalement dans ce groupe des cours sur les activités d'éveil à l'éducation préscolaire (PPE32203) et la gestion éducative de la classe au primaire (PPE36002).

**Tableau 4. Coefficients factoriels standardisés moyens (bases de données 1 à 5)**

| Sigle    | 1    | 2    | 3    | 4    | R <sup>2</sup> |
|----------|------|------|------|------|----------------|
| ATE40002 |      | 0,68 |      |      | 0,50           |
| DIA13002 |      |      | 0,77 |      | 0,66           |
| DIA15003 | 0,29 | 0,30 | 0,44 |      | 0,40           |
| DIA23013 |      | 0,50 | 0,36 | 0,47 | 0,60           |
| DIA24002 |      |      | 0,39 |      | 0,32           |
| DIA25003 | 0,58 |      | 0,37 |      | 0,52           |
| DIA28202 | 0,75 |      |      |      | 0,60           |
| DIA33003 |      | 0,72 | 0,39 |      | 0,71           |
| DIA35003 | 0,63 | 0,33 |      |      | 0,55           |
| DIA43003 |      | 0,58 |      | 0,42 | 0,56           |
| DIA45002 | 0,72 |      |      |      | 0,55           |
| PPE10002 |      |      |      | 0,61 | 0,47           |
| PPE11002 | 0,33 |      | 0,50 | 0,39 | 0,57           |
| PPE11102 | 0,54 |      | 0,30 | 0,35 | 0,52           |
| PPE18002 | 0,69 |      |      |      | 0,57           |
| PPE22002 |      | 0,57 |      | 0,38 | 0,53           |
| PPE22102 |      | 0,62 | 0,39 |      | 0,59           |
| PPE28102 | 0,38 | 0,36 | 0,31 |      | 0,44           |
| PPE32203 | 0,58 |      |      | 0,34 | 0,46           |
| PPE36002 | 0,47 |      |      | 0,54 | 0,53           |
| PPE37002 | 0,38 | 0,46 |      | 0,38 | 0,52           |
| PPE40012 |      | 0,56 | 0,34 |      | 0,47           |
| SCE10002 |      |      | 0,35 | 0,41 | 0,33           |
| SCE10102 |      |      | 0,39 |      | 0,30           |
| SCE20002 | 0,32 |      | 0,61 |      | 0,56           |
| SCE20202 |      |      | 0,74 |      | 0,62           |
| SCE20303 |      |      |      | 0,71 | 0,59           |
| SCE30102 |      | 0,32 |      | 0,56 | 0,45           |

53 Le second facteur regroupe des cours qui gravitent autour de la didactique du français et des fondements de l'éducation. Il regroupe des cours de didactique en français et univers social (DIA23013, DIA33003 et DIA43003). Le volet préscolaire est très présent dans ce groupe de didactique avec les cours PPE22002 et PPE22102. On retrouve également des cours d'approfondissement d'un thème en éducation (ATE40002), de projet personnel de formation :

synthèse (PPE40012) et d'évaluation (PPE37002 et SCE30102). Des cours de didactique des mathématiques (DIA15003 et DIA35003) et de fondements de l'éducation (PPE28102) saturent également dans ce facteur.

54 Le troisième facteur est l'adaptation de l'enseignement. Il est fortement associé aux cours de prévention, d'adaptation et de réussite scolaire (SCE20202) et de connaissance et d'intégration des élèves handicapés et en difficulté (SCE20002). Ce facteur d'adaptation de l'enseignement rejoint plusieurs autres cours de didactique du français (DIA13002, DIA23013 et DIA33003), des mathématiques (DIA15003 et DIA25003) ainsi que de la géographie, histoire et éducation à la citoyenneté (DIA24002). Certains cours en fondement de l'éducation s'y retrouvent également tels que le développement de l'enfant (PPE11002 et PPE11102), l'utilisation pédagogique des technologies de l'information et de la communication (SCE10102), et le projet personnel de formation : synthèse (PPE40012).

55 Le quatrième facteur est associé au développement des aspects personnels et sociaux. Tous les cours d'éducation préscolaire s'insèrent dans ce facteur (PPE36002, PPE22002 et PPE32203), ainsi que des cours tels que le projet personnel de formation : élaboration (PPE10002), l'organisation scolaire (SCE10002) et le développement de l'enfant (PPE11002 et PPE11102). Les cours en évaluation (SCE30102 et PPE37002) s'y retrouvent ainsi que le cours d'interventions auprès des élèves en trouble du comportement (SCE20303). Des cours de didactique du français (DIA23013 et DIA43003) s'insèrent aussi dans ce facteur.

### 3.5. Étape 5 : l'étude de la fiabilité

56 La fiabilité de l'ensemble des cours est forte et est estimée à 0,92 par le coefficient alpha de Cronbach sur la base de données initiale. Il n'y a pas de différence entre les coefficients alpha de Cronbach générés à partir des autres bases de données. Le critère fixé à 0,60 est atteint.

### 3.6. Étape 6 : la validité de critère

57 A priori, l'analyse montre qu'il existe des corrélations importantes entre la cote de rendement au collège des étudiants et la moyenne cumulative à l'université pour le programme considéré ( $r = 0,69$ ,  $p < 0,001$ ). Il existe des corrélations moindres entre la moyenne cumulative à l'université et le campus d'études ( $r = 0,13$ ,  $p < 0,001$ ), ainsi qu'avec le genre ( $r = -0,24$ ,  $p < 0,001$ ) et l'année où le diplôme a été obtenu ( $r = -0,11$ ,  $p < 0,001$ ), mais pas avec l'âge du diplômé ( $r = -0,05$ ,  $p = 0,17$ ).

58 Afin de clarifier l'apport respectif de chacune de ces variables en tenant compte de la présence des autres, des analyses de régression à multiples niveaux sont réalisées (voir équations 1.0 à 5.0 et tableau 5). L'étude du modèle inconditionnel montre que 97,7 % de la variance s'explique par des différences intragroupes (0,13/0,133). Le modèle 1 permet d'expliquer les différences individuelles (intragroupe) au regard de la cote de rendement au collège. L'ajout de cette variable réduit la variance intragroupe de 53,84 % (soit  $1 - 0,06/0,13$ ). La cote de rendement au collège est donc une excellente variable afin de prédire le succès à l'université d'un futur étudiant dans le programme en éducation préscolaire et en enseignement primaire. Le critère visé, soit 30 % de variance expliquée, est atteint et le pourcentage de variance expliquée est considéré équivalent à ce qui est rapporté pour d'autres programmes (Brière, 1998).

### 3.7. Étape 7 : l'équité entre les étudiants

59 Le genre et l'âge ont une influence, mais faible (voir modèle 2). Ainsi, ces deux variables combinées ne permettent qu'une réduction de 2,3 % de la variance individuelle (soit de 0,06 à 0,057). Cet apport est cependant significatif. Comme il a été mentionné précédemment, l'effet de l'âge n'est pas présent lorsque l'échantillon total est considéré et que s'ajoutent les candidats admis sur la base d'un diplôme équivalent ou sur la base adulte. Le critère visé, qui stipule qu'aucune différence en fonction de l'âge et du genre ne devrait exister, n'est pas atteint. Il existe certaines iniquités en fonction de l'âge et du genre de l'étudiant.

### 3.8. Étape 8 : l'équivalence entre les groupes d'étudiants

60 Les effets « Campus » et « Année » sont non-significatifs. Ils n'expliquent respectivement que 1,50 % (soit 0,002/0,133) et 0,75 % (soit 0,001/0,133) de la variance totale (voir le modèle inconditionnel). Il n'y a donc pas d'inflation des notes entre les cohortes d'étudiants diplômés de 2006 à 2012. Le critère visé, qui stipule qu'aucune différence en fonction du campus et de l'année d'études ne devrait exister, est atteint<sup>9</sup>.

**Tableau 5. Régressions à niveaux multiples**

| Effets fixes                   | Modèles multiniveaux       |          |          |
|--------------------------------|----------------------------|----------|----------|
|                                | Inconditionnel             | Modèle 1 | Modèle 2 |
| Constante ( $\theta$ )         | 3,39*                      | 1,32*    | 0,84*    |
| CRC ( $\Pi_1$ )                |                            | 0,079*   | 0,08*    |
| Genre <sup>1</sup> ( $\Pi_2$ ) |                            |          | 0,19*    |
| Âge <sup>2</sup> ( $\Pi_3$ )   |                            |          | 0,02*    |
| Effets aléatoires              | Composantes de la variance |          |          |
| Intragroupe ( $\sigma^2$ )     | 0,13*                      | 0,06*    | 0,057*   |
| Campus ( $\square_{b00}$ )     | 0,002                      | 0,005    | 0,002    |
| Année ( $\square_{c00}$ )      | 0,001                      | 0,004    | 0,004    |
| Déviance                       | 714,9                      | 81,4     | -1,8     |
| BIC                            | 714,9                      | 81,4     | -1,8     |

<sup>1</sup> Différence en faveur des diplômées féminines

<sup>2</sup> Différence en faveur des étudiants plus âgés

## 4. Discussion

61 À ce stade-ci, il est nécessaire de discuter de la validité de construit curriculaire, de la validité de critère, de la fiabilité, de l'équité entre les individus ainsi que de l'équivalence entre les groupes.

62 En ce qui a trait à la validité de construit curriculaire, les notes obtenues dans les cours du programme analysé sont corrélées et révèlent la présence de quatre groupes de facteurs qui expliquent plus de 50 % de la variance totale. La fiabilité de l'ensemble des cours est excellente et se situe au-delà de 90 %. En ce sens, le programme respecte les principaux critères de validité et de fiabilité fixés par Rexwinkel, Haenen et Pilot (2013). Par conséquent, la méthodologie présente une valeur diagnostique. Elle permet aux équipes professorales concernées d'effectuer une réflexion sur la contribution de chacun des cours à la mesure de construits communs. Les construits obtenus sont-ils ceux attendus au terme du programme ? Quelle est alors la contribution (saturation) de chacun des cours à ces construits ?

63 La principale difficulté réside dans l'interprétation des facteurs après rotation en lien avec les prescriptions du référentiel. Celui-ci suggère que les compétences visées se regroupent en quatre construits liés aux fondements de l'éducation, à l'acte d'enseigner, aux aspects sociaux et au développement identitaire (MEQ, 2001). Or, les quatre construits qui émergent des analyses sont spécifiques aux fondements de l'éducation liés à l'acquisition des compétences en mathématiques, aux fondements de l'éducation liés à l'acquisition des compétences en français, à l'acquisition des aspects personnels et sociaux, et à l'adaptation de l'acte d'enseigner. Il n'y a donc pas une adéquation parfaite entre les facteurs obtenus et les groupes de compétences prescrits par le référentiel. Notamment, la présence d'une composante liée à

- l'acquisition des compétences en mathématiques ne permet pas l'utilisation d'une approche confirmatoire où la structure factorielle attendue au terme du programme est utilisée afin d'expliquer la structure de la matrice de corrélation entre les cours (Harvey, 2012).
- 64 Difficile de dire à ce stade-ci s'il faut modifier le référentiel ou revoir la pratique évaluative exercée dans le cadre du programme. L'analyse d'autres programmes d'éducation préscolaire et d'enseignement primaire du Québec sera nécessaire pour déterminer si des construits spécifiques associés aux compétences en didactique des mathématiques et en didactique du français émergent ou s'il s'agit d'une particularité du programme analysé. La présence de compétences distinctes en mathématiques et en langue a déjà été rapportée chez les étudiants du secondaire en Suède à partir de l'analyse des résultats dans les cours spécifiques à ces matières et aux tests nationaux (Lekholm & Cliffordson, 2008 ; Thorsen & Cliffordson, 2012). Ces travaux justifient que de telles compétences spécifiques puissent être évaluées dans le cadre du programme d'enseignement au primaire, car ces matières font partie du curriculum obligatoire au primaire.
- 65 Une autre difficulté est que l'analyse ne permet pas de distinguer si les construits évalués sont des groupes de compétences ou des ressources (par exemple, connaissances) liées à l'acquisition de celles-ci. L'analyse des résultats académiques ne permet pas d'effectuer cette distinction ; des indicateurs supplémentaires (Mason & Dragovich, 2010) ou des preuves issues d'une analyse du contenu des évaluations sont nécessaires pour effectuer cette inférence.
- 66 Par ailleurs, la validité de critère des dossiers universitaires est excellente. Les différences individuelles dans la moyenne cumulative s'expliquent en grande partie par des différences dans les connaissances/compétences antérieures de l'étudiant (CRC). Cette capacité des résultats antérieurs à prédire les succès universitaires rejoint les écrits sur le sujet (Camara & Echternacht, 2000 ; DiPerna, 2004 ; Geiser & Santelices, 2007 ; Pattison, Grodsky & Muller, 2013 ; Pfister, 1975 ; Kobrin *et al.*, 2008 ; Komarraju, Ramsay & Rinella, 2013 ; Pike & Saupe, 2002). Il est cependant important de constater que cette relation entre les succès académiques antérieurs et actuels des étudiants est réitérée dans le cadre du programme analysé. La fiabilité des notes à l'ensemble des cours est également satisfaisante et rejoint ce qui a été observé ailleurs (Harvey, 2012 ; Saupe & Eimers, 2012).
- 67 L'étude de l'équité entre les individus est, dans son ensemble, positive. Transparaissent cependant, de légères iniquités entre le genre et l'âge des étudiants, et ce, après que l'effet des résultats antérieurs soit considéré. Les diplômées obtiennent des notes supérieures à celles des diplômés. Un avantage de la même magnitude (2 %) en faveur des filles est également rapporté par Rojstaczer et Healy (2012). De même, les étudiants plus âgés obtiennent de meilleurs résultats. L'effet combiné de ces deux facteurs sur la moyenne cumulative est cependant faible. Tel que mentionné antérieurement, une explication possible est que les étudiants plus âgés adoptent des comportements plus matures ; ces comportements sont alors favorablement notés par les ressources professorales en contexte de classe. Il faut être prudent dans l'interprétation de ce dernier résultat, car les étudiants admis sur une base autre que le diplôme d'études au collège sont exclus de ces analyses.
- 68 L'équivalence entre les deux campus et entre les années de formation fut également investiguée. Ainsi, malgré la présence de ressources différentes (par exemple, professeurs) dans les deux campus où le programme est offert, les résultats académiques obtenus par les étudiants sont équivalents. Comme il s'agit de la même institution universitaire, il est possible de considérer que le contexte et climat des deux campus sont relativement équivalents et que les différences rapportées ailleurs en lien avec ces facteurs (Ma *et al.*, 2008) ne s'appliquent pas dans la présente étude. De même, il n'y a pas de différence entre les deux campus quant aux facteurs (public/privé, nord/sud, technique/non-technique) qui influencent les résultats académiques tels que rapportés par Rojstaczer et Healy (2012). Ici encore, une certaine prudence est requise dans l'interprétation de ces données, car le nombre de groupes de second niveau est relativement faible. Selon Maas et Hox (2005), les estimés de la variance de second niveau peuvent être sous-estimés d'environ 15 % lorsque le nombre de groupes est inférieur à 50.

69 De même, il n'y a pas, dans la période considérée, de fluctuations attribuables à l'année de diplomation ; le processus d'inflation des notes rapporté dans les écrits (Caruth & Caruth, 2013 ; Jewell & McPherson, 2012 ; Rojstaczer & Healy, 2012) semble stabilisé, du moins en ce qui concerne le programme et les années considérées. La moyenne observée est, en soi, relativement élevée (B+) ; elle s'avère supérieure à la notation « C » qui représente la norme en vigueur avant le phénomène rapporté d'inflation des notes. Cependant, comme les résultats académiques ont démontré une structuration intéressante en matière de construits mesurés, de validité de critère et de fiabilité, cette moyenne ne constitue pas une menace en soi (Millman *et al.*, 1983 ; Pattison, Grodsky & Muller, 2013) au signal envoyé à la société quant à la qualité des notes accordées aux diplômés.

## 5. Conclusion

70 L'objectif poursuivi par cette étude permet d'informer et éclairer les acteurs sociaux quant aux qualités des résultats académiques d'un programme de formation en éducation préscolaire et en enseignement primaire d'une université francophone. L'analyse possède une valeur diagnostique, car elle permet de vérifier la qualité des notes académiques et la contribution de chacun des cours à ces qualités. Il s'agit d'éléments importants dans le cadre d'une évaluation considérée socialement et éthiquement légitime (Hadji, 2012), qui alimente la réflexion de ces mêmes acteurs.

71 Ainsi, l'étude a permis de vérifier la validité de construit curriculaire du programme, sa validité de critère ainsi que la fiabilité des évaluations réalisées. Une attention particulière a été portée à l'équité entre les individus en fonction du genre et de l'âge des diplômés ainsi qu'à la validité de construit curriculaire, la validité de critère, la fiabilité et l'équivalence entre des groupes ayant étudié à des campus différents et diplômé à des moments différents. Ceux-ci ont notamment partagé des ressources différentes dans l'espace géographique et dans le temps. Malgré la présence de quelques iniquités en regard du genre et de l'âge des diplômés, les résultats académiques du programme analysé démontrent de belles qualités en ce qui concerne l'équivalence entre des groupes d'étudiants.

72 Par l'ajout de questions relatives à l'équité entre les individus, à l'équivalence entre les groupes et à la validité de critère, la présente méthodologie va au-delà de celle proposée par Harvey (2012) et Rexwinkel, Haenen et Pilot (2013) ; en cela elle représente une contribution originale aux écrits.

---

### Bibliographie

- Agazzi, A. (1967). *Les aspects pédagogiques des examens*. Strasbourg : Conseil de l'Europe.
- American Educational Research Association (AERA)(1999). *Standards for educational and psychological testing*. Washington, DC : American Educational Research Association.
- Ayotte-Beaudet, J.-P. (2013). The concept of competence in the French-language education literature. *Prospect*, 168, 419-427.
- Bacon, D.R. & Bean, B. (2006). GPA in research studies : An invaluable but neglected opportunity. *Journal of Marketing Education*, 28(1), 35-42. doi : 10.1177/0273475305284638.
- Beauvais, M. (2012/3). Des éthiques de la responsabilité pour penser l'évaluation en éducation et en formation. *Empan*, 87, 132-139. doi : 10.3917/empan.087.0132.
- Beretvas, S.N. (2008). Cross-classified random effects models. In A.A. O'Connell & D.B. McCoach (Eds), *Multilevel modeling of educational data* (pp. 161-197). Charlotte, NC : Information Age Publishing.
- Brière, L. (1998). *La cote de rendement au collégial (CRC) et la prédiction du succès universitaire en psychologie, en administration et en enseignement de l'activité physique (Mémoire de maîtrise inédit)*. Université du Québec à Trois-Rivières.
- Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. New York, NY : Guilford Press.
- Buysse, A.A.J. (2012). Portfolio de développement professionnel et certification : Frictions constructives ou destructives. *Le bulletin de l'ADMEE-Europe*, 2012(1), 25-39.

- Camara, W.J. & Echternacht, G. (2000). *The SAT I and high school grades : Utility in predicting success in college* (Research Notes RN-10). New York, NY : The College Board, Office of Research and Development.
- Caruth, D.L. & Caruth, G.D. (2013). Grade inflation : An issue for higher education ? *Turkish Online Journal of Distance Education*, 14(1), 102-110.
- Council of Higher Education Accreditation (2003). *Statement of mutual responsibilities for student learning outcomes : Accreditation, institutions, and programs*. Washington, DC : Council for Higher Education Accreditation.
- Collège d'enseignement général et professionnel Marie-Victorin (2005). *Politique no-12, Politique institutionnelle d'évaluation des apprentissages*. Repéré à [http://www.collegemv.qc.ca/CMS/Media/2226\\_294\\_fr-CA\\_0\\_pol\\_12\\_PIEA.pdf](http://www.collegemv.qc.ca/CMS/Media/2226_294_fr-CA_0_pol_12_PIEA.pdf).
- Comité d'Agrément des Programmes de Formation à l'Enseignement (CAPFE) (2010). *La visite de suivi de l'agrément d'un programme de formation à l'enseignement : cadre de référence et modalités d'application pour les visites de suivi de l'agrément qui seront effectuées entre 2010 et 2013*. Québec, QC : Ministère de l'Éducation, du Loisir et du Sport.
- Conseil supérieur de l'éducation (CSE) (2012). *L'assurance qualité à l'enseignement universitaire : une conception à promouvoir et à mettre en œuvre. Avis à la ministre de l'Éducation, du Loisir et du Sport*. Québec, QC : Gouvernement du Québec.
- DiPerna, J.C. (2004). Structural and concurrent validity evidence for the Academic Competence Evaluation Scales-College edition. *Journal of College Counseling*, 7, 64-72. doi : 10.1002/j.2161-1882.2004.tb00260.x.
- European Network for Quality Assurance in Higher Education (ENQA) (2009). *Standards and guidelines for quality assurance in the European higher education area*. Helsinki, Finland :ENQA.
- Geiser, S. & Santelices, M.V. (2007). *Validity of high-school grades in predicting student success beyond the freshman year : High-school record vs. standardized tests as indicators of four-year college outcomes*. Research & Occasional Paper Series : CSHE.6.07, University of California at Berkeley.
- Goova, M.T., Hollett, L.A., Tesfay, S.T., Gala, R.B., Puzziferri, N., Kehdy, F.J. & Scott, D.J. (2008). Implementation, construct validity, and benefit of a proficiency-based knot-tying and suturing curriculum. *Journal of Surgical Education*, 65(4), 309-315. doi : 10.1016/j.jsurg.2008.04.004.
- Hadji, C. (2012). *Faut-il avoir peur de l'évaluation ?* Bruxelles : De Boeck.
- Hango, D. (2013). Gender differences in science, technology, engineering, mathematics and computer science (STEM) programs at university. *Insights on Canadian Society, December*, Statistics Canada Catalogue no. 75-006-X.
- Harvey, L.(2012). Évaluation des compétences dans un programme de formation en enseignement : Validité de construit curriculaire. *Mesure et évaluation en éducation*, 35(2), 69-97.
- Harvey, L., Hébert, M.-H. & Simard, C. (2013). Qualités métriques des dossiers académiques universitaires. *Revue canadienne de l'éducation*, 36(3), 160-183.
- Jabor, M.K., Machtmes, K., Kungu, K., Buntat, Y. & Nordin, M. S. (2011). The influence of age and gender on the student achievement in mathematics. Paper presented at the 2011 International Conference on Social Science and Humanity, IPEDR vol. 5. (2011). Singapour : IACSIT Press. Repéré à <http://www.ipedr.com/vol5/no2/67-H10178.pdf>.
- Jewell, R.T. & McPherson, M. A. (2012). Instructor-specific grade inflation : Incentives, gender, and ethnicity. *Social Science Quarterly*, 93(1), 95-109. doi : 10.1111/j.1540-6237.2011.00827.x.
- Joint Committee on Standards for Educational Evaluation (JCSEE) (2012). *Classroom Assessment Standards Draft 5*. Repéré à <http://www.jcsee.org/standards-development>.
- Kobrin, J.L., Patterson, B.F., Shaw, E.J., Mattern, K.D. & Barbuti, S.M. (2008). *Validity of the SAT for predicting first-year college grade point average* (Research report no-2008-5). New York, NY : The College Board.
- Komaraju, M., Ramsey, A. & Rinella, V. (2013). Cognitive and non-cognitive predictors of college readiness and performance : Role of academic discipline. *Learning and Individual Differences*, 24, 103-109. doi : 10.1016/j.lindif.2012.12.007.
- Kuncel, N.R., Credé, M. & Thomas, L. L. (2007). A meta-analysis of the predictive validity of the graduate management admission test (GMAT) and undergraduate grade point average (UGPA) for graduate student academic performance. *Academy of Management Learning & Education*, 6(1), 51-68. doi : 10.5465/AMLE.2007.24401702.



- Kuncel, N.R., Wee, S., Serafin, L. & Hezlett, S.A. (2010). The validity of the graduate record examination for master's and doctoral programs : A meta-analytic investigation. *Educational and Psychological Measurement*, 70(2), 340-352. doi : 10.1177/0013164409344508.
- Lekholm, A.K. & Cliffordson, C. (2008). Discrepancies between school grades and test scores at individual and school level : Effects of gender and family background. *Educational Research and Evaluation*, 14(2), 181-199. doi : 10.1080/13803610801956663.
- Lewis, R. (2011). L'avenir de l'assurance-qualité au sein du système mondial d'enseignement supérieur. Dans OCDE (Ed.), *L'enseignement supérieur à l'horizon 2030 – Volume 2 : mondialisation, La recherche et l'innovation dans l'enseignement*. Repéré à <http://www.oecd-ilibrary.org/content/book/9789264075405-fr>.
- Luthy, T.L. (1996). *Validity and prediction bias of grade performance from Graduate Record Examination scores for students at Northern Illinois University : Age and gender considerations*. Northern Illinois University, ProQuest, UMI Dissertation Publishing.
- Ma, X., Ma, L. & Bradley, K.D. (2008). Using multilevel modeling to investigate school effects. In A.A. O'Connell & D.B. McCoach (Eds), *Multilevel modeling of educational data* (pp. 59-110). Charlotte, NC : Information Age Publishing.
- Maas, C.J.M. & Hox, J.J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1(3), 86-92. doi : 10.1027/1614-1881.1.3.86.
- Mason, G. & Dragovich, J. (2010). Program assessment and evaluation using student grades obtained on outcome-related course learning objectives. *Journal of Professional Issues in Engineering Education and Practice*, 136(4), 206-214. doi : 10.1061/(ASCE)EI.1943-5541.0000029.
- McCoach, D.B. & Black, A.C. (2008). Evaluation of model fit and adequacy. In A. A. O'Connell & D. B. McCoach (Eds), *Multilevel modeling of educational data* (pp. 245-273). Charlotte, NC : Information Age Publishing.
- Millman, J., Slovacek, S.P., Kulick, E. & Mitchell, K.J. (1983). Does grade inflation affect the reliability of grades ? *Research in Higher Education*, 19(4), 423-429. doi : [10.1007/BF01418444](https://doi.org/10.1007/BF01418444).
- Ministère de l'Éducation du Québec (MEQ) (2001). *La formation à l'enseignement : les orientations, les compétences professionnelles*. Québec, QC : Gouvernement du Québec.
- Ministère de l'Éducation du Québec (MEQ) (2003). *Politique d'évaluation des apprentissages*. Repéré à [http://www.mels.gouv.qc.ca/fileadmin/site\\_web/documents/publications/EPEPS/Formation\\_jeunes/Evaluation/13-4602.pdf](http://www.mels.gouv.qc.ca/fileadmin/site_web/documents/publications/EPEPS/Formation_jeunes/Evaluation/13-4602.pdf).
- Pattison, E., Grodsky, E. & Muller, C. (2013). Is the sky falling ? Grade inflation and the signaling power of grades. *Educational Researcher*, 42(5), 259-265. doi : 10.3102/0013189X13481382.
- Pfister, C. (1975). *La validité de la note scolaire (Thèse de doctorat inédite)*. Université de Neuchâtel.
- Pike, G.R. & Saupe, J.L. (2002). Does high school matter ? An analysis of three methods of predicting first-year grades. *Research in Higher Education*, 43(2), 187-207.
- Prégent, R., Bernard, H. & Kozanitis, A. (2009). *Enseigner à l'université dans une approche-programme : un défi à relever*. Montréal, QC : Presses internationales Polytechnique.
- Rexwinkel, T., Haenen, J. & Pilot, A. (2013). Quality assurance in higher education : Analysis of grades for reviewing course levels. *Quality and Quantity*, 47(1), 581-598. doi : 10.1007/s11135-011-9481-6.
- Rey, B. (2008). Quelques aspects éthiques de l'évaluation. Dans Baillat Gilles, De Ketele Jean-Marie, Paquay Léopold & Thélot Claude (dir.). *Évaluer pour former. Outils, dispositifs et acteurs* (pp. 57-67). Bruxelles : De Boeck.
- Rojstaczer, S. & Healy, C. (2012). Where A is ordinary : The evolution of american college and university grading, 1940-2009. *Teachers College Record*, 114(7), 1-23.
- Ryan, K.E. & Cousins, J.B. (2009). Introduction. In K.E. Ryan & J.B. Cousins (Eds), *The SAGE international handbook of educational evaluation* (pp. ix-xvii). Los Angeles, CA : SAGE.
- Saupe, J.L. & Eimers, M.T. (2012). *Alternative estimates of the reliability of college grade point averages*. Annual Forum of the Association for Institutional Research, June 2 - June 6, 2012, New Orleans, Louisiana.
- Smith, D.L. (1992). Validity of faculty judgments of student performance : Relationship between grades and credits earned and external criterion measures. *The Journal of Higher Education*, 63(3), 329-340. doi : [10.2307/1982018](https://doi.org/10.2307/1982018).

Thorsen, C. & Cliffordson, C. (2012). Teachers' grade assignment and the predictive validity of criterion-referenced grades. *Educational Research and Evaluation : An International Journal on Theory and Practice*, 18(2), 153-172.

---

### Notes

1 Les expressions « dossier académique », « dossier universitaire », « relevé de notes » et « notes académiques » sont considérées équivalentes dans le présent texte.

2 La proportion de femmes est respectivement de 92,78 % et 92,33 % aux campus de Lévis et Rimouski. Cette différence est non significative ( $\chi^2 = 0,06$ ,  $df = 1$ ,  $p = 0,80$ ). L'âge des diplômés est respectivement de 24,17 et 24,65 aux campus de Lévis et Rimouski avec des écarts-types de 2,48 et 3,77. Cette différence de 0,48 est équivalente à une différence de 6 mois, qui est considérée significative ( $t = -2,22$ ,  $p = 0,03$ ).

3 En contexte de cours offerts en institution, les savoirs professionnels sous-jacents aux compétences sont souvent évalués plutôt que les compétences elles-mêmes (Buisse, 2012).

4 Afin de rendre ce critère comparable avec d'autres échelles de mesure, il sera traduit en coefficient de variation ( $\chi/\#$ ). Le critère de Rexwinkel Haenen et Pilot (2013) suggère donc que des coefficients de variation d'environ 10 % sont acceptables mais que des coefficients de 4 % et 20 % sont respectivement trop faibles et trop élevés.

5 La logique s'inspire de ce qui est présenté par Brown (2006). Elle s'applique à la situation où un seul construit latent est présent, mais celle-ci peut être facilement généralisée à la situation où plusieurs construits latents sont évalués. Ainsi, la note académique ( $y$ ) attribuée à un cours devient une fonction du construit latent ( $\#$ ) et de la composante propre à chaque indicateur ( $\#$ ). Cette composante propre est aussi parfois interprétée comme étant l'erreur de mesure. L'équation suivante exprime cette fonction,

6 Les notes littérales [A+...E] utilisées pour chacun des cours furent converties en cotes numériques en utilisant le barème suivant : « A+ # 4,3 », « A # 4,0 », « A- # 3,7 » et ainsi de suite jusqu'à « E # 0 ». L'échelle obtenue compte 12 niveaux et est généralement considérée et traitée comme une échelle d'intervalle.

7 Des coefficients de variation se situant entre 12,5 % et 31,8 % sont observés dans notre étude. Plusieurs coefficients dépassent le 20 % inféré de Rexwinkel, Haenen et Pilot (2013). Ces grandes variations s'expliquent par le grand nombre d'étudiants et leur diversité, la durée de la période échantillonnée, la diversité des ressources professorales etc. À ce stade-ci des analyses, cette grande dispersion n'est pas en soi préoccupante.

8 Considérant sa grande taille, cette matrice n'est pas publiée ici, mais peut être obtenue sur demande auprès des auteurs.

9 La normalité de distribution des erreurs ( $e_{i(j),j2}$ , voir équation 1.0) a été vérifiée à l'aide d'un test du Khi-carré. Celui-ci s'est avéré non-significatif ( $\chi^2 = 12,5$ ,  $df = 10$ ,  $p = 0,25$ ) ; ce qui suggère que les erreurs sont normalement distribuées, tel que le présume ce modèle.

---

### Pour citer cet article

#### Référence électronique

Léon Harvey, Marie-Hélène Hébert et Catherine Simard, « Qualités des dossiers universitaires : analyse d'un programme de formation en enseignement », *Revue internationale de pédagogie de l'enseignement supérieur* [En ligne], 30-3 | 2014, mis en ligne le 01 décembre 2014, consulté le 09 février 2016. URL : <http://ripes.revues.org/884>

---

### À propos des auteurs

#### Léon Harvey

Unité des sciences de l'éducation, Université du Québec à Rimouski  
300, allée des Ursulines, Rimouski (Québec) G5L 3A1  
[Leon\\_Harvey@uqar.ca](mailto:Leon_Harvey@uqar.ca)

#### Marie-Hélène Hébert

Unité des sciences de l'éducation, Université du Québec à Rimouski  
300, allée des Ursulines, Rimouski (Québec) G5L 3A1  
[Marie-Helene\\_Hebert@uqar.ca](mailto:Marie-Helene_Hebert@uqar.ca)

#### Catherine Simard

Unité des sciences de l'éducation, Université du Québec à Rimouski

300, allée des Ursulines, Rimouski (Québec) G5L 3A1  
Catherine\_Simard@uqar.ca

---

### ***Droits d'auteur***

Article L.111-1 du Code de la propriété intellectuelle.

---

### ***Résumés***

Cette étude présente une lecture des dossiers académiques d'étudiants ayant complété une formation universitaire en éducation préscolaire et en enseignement primaire afin de vérifier la validité et la fiabilité des évaluations réalisées durant cette formation. L'importance de facteurs tels que la cote de rendement au collège des étudiants, l'équivalence entre deux campus universitaires et les années de diplomation, ainsi que l'équité entre le genre et l'âge des étudiants est également investiguée. La base de données se compose des dossiers académiques de 873 finissants du programme d'éducation préscolaire et d'enseignement primaire ayant gradué entre 2006 et 2012 dans l'un ou l'autre des deux campus de l'Université du Québec à Rimouski. Les résultats de l'analyse factorielle suggèrent que les notes ne sont pas le produit d'un processus arbitraire et aléatoire. Ainsi, l'ensemble des notes aux cours du programme, telles qu'elles apparaissent dans le dossier académique des étudiants, se structure en facteurs qui permettent d'inférer, tel qu'attendu, la mesure de groupes de compétences interreliées. L'ensemble possède une bonne fiabilité. Des régressions multiniveaux montrent que la cote de rendement au collège est un bon prédicteur de la moyenne cumulative et qu'il y a équivalence en fonction du campus d'études et de l'année de diplomation. Cependant, il subsiste de légères iniquités dans la moyenne cumulative en fonction de l'âge et du genre de l'étudiant, iniquités qui ne sont pas liées aux succès académiques antérieurs.

This study presents a reading of academic files of students having completed a bachelor degree in a preschool education and primary teaching curriculum in order to assess the validity and reliability of the underlying grades. The impact of factors such as previous college results, campus of study, year of diploma, gender and age are also investigated. The database is composed of 873 graduated students from the preschool education and primary teaching program, having graduated between 2006 and 2012 from one of the two campuses of the University of Quebec at Rimouski. The results of an exploratory factorial analysis suggest that the grades are not the product of random and arbitrary processes. Student grades according to their records are structured according to factors that, as expected, help measure four groups of connected learning outcomes. The set shows good reliability. Multiple level regressions reveal that past college performance is a good predictor of university mean grade point average (UGPA) with no impact of the university campus or the year of diploma. However, some slight iniquities remain in the UGPA as a function of students' gender and age, once past academic success has been accounted for.

### ***Entrées d'index***

***Index de mots-clés*** : équité, fiabilité, programme de formation, résultats académiques, validité