

# Les évaluations de l'enseignement par les étudiants : vers une démarche abrégée

Léon Harvey

Volume 39, numéro 2, 2016

Réception : 6 juillet 2015

Version finale : 7 juin 2016

Acceptation : 9 juin 2016

URI : <https://id.erudit.org/iderudit/1038240ar>

DOI : <https://doi.org/10.7202/1038240ar>

[Aller au sommaire du numéro](#)

### Éditeur(s)

ADMEE-Canada - Université Laval

### ISSN

0823-3993 (imprimé)

2368-2000 (numérique)

[Découvrir la revue](#)

### Citer cet article

Harvey, L. (2016). Les évaluations de l'enseignement par les étudiants : vers une démarche abrégée. *Mesure et évaluation en éducation*, 39(2), 1-37.  
<https://doi.org/10.7202/1038240ar>

### Résumé de l'article

Cet article documente le processus de simplification de la démarche d'évaluation de l'enseignement par les étudiants. Cette simplification repose sur la conception d'un instrument abrégé qui permet de porter des appréciations valides et fiables sur la qualité de l'enseignement. Une comparaison des qualités psychométriques entre l'instrument original et abrégé est ainsi réalisée. En conclusion, certaines modifications de la version abrégée sont proposées, mais cette version représente néanmoins un substitut adéquat à la version originale.

## Les évaluations de l'enseignement par les étudiants : vers une démarche abrégée

Léon Harvey

Université du Québec à Rimouski

**MOTS CLÉS:** évaluation, enseignement, étudiants, validité, comportement séquentiel, fiabilité, questionnaire abrégé

*Cet article documente le processus de simplification de la démarche d'évaluation de l'enseignement par les étudiants. Cette simplification repose sur la conception d'un instrument abrégé qui permet de porter des appréciations valides et fiables sur la qualité de l'enseignement. Une comparaison des qualités psychométriques entre l'instrument original et abrégé est ainsi réalisée. En conclusion, certaines modifications de la version abrégée sont proposées, mais cette version représente néanmoins un substitut adéquat à la version originale.*

**KEY WORDS:** evaluation, teaching, students, validity, sequential behavior, reliability, short questionnaire

*This paper documents the process of simplifying the student course evaluation procedure. The simplification relies on the conception of a short questionnaire that supports the enunciation of valid and reliable judgments about the quality of a course. For such, a comparison of the psychometric qualities of the original and short versions is realized. To conclude, some modifications of the short version are proposed, but it nevertheless represents an adequate substitute to the original version.*

PALAVRAS-CHAVE: avaliação, ensino, estudantes, validade, comportamento sequencial, fiabilidade, questionário abreviado

*Este artigo documenta o processo de simplificação do procedimento de avaliação do ensino por via dos estudantes. Esta simplificação baseia-se na conceção de um instrumento abreviado que permite realizar apreciações válidas e fiáveis sobre a qualidade do ensino. Assim, alcança-se uma comparação das qualidades psicométricas entre o instrumento original e o abreviado. Finalmente, são propostas certas alterações à versão abreviada, mas esta versão, não obstante, representa um substituto adequado da versão original.*

## Introduction

Cet article documente le processus de simplification de la démarche d'évaluation de l'enseignement par les étudiants. Pour ce faire, nous procéderons à l'étude des qualités psychométriques des versions originale et abrégée d'un questionnaire d'évaluation de l'enseignement par les étudiants (EEE). À l'instar de Dolmans et Ginns (2005), cet article considère qu'un questionnaire abrégé d'évaluation de l'enseignement peut être valide et fiable. Le format de l'instrument utilisé à l'université ciblée est hybride en ce qu'il se compose de questions fermées qui s'accompagnent d'une section ouverte permettant aux étudiants d'exprimer dans leurs propres mots leur appréciation du cours qu'ils ont suivi. Il diffère ainsi des questionnaires fermés (p. ex., Marsh & Bailey, 1993) ou ouverts généralement rapportés dans la littérature. L'impact d'une version abrégée sur la capacité des étudiantes et étudiants à exprimer leur opinion au moyen de commentaires écrits sera donc également documenté. De plus, la version abrégée modifie la structure séquentielle des choix de réponse. Il s'avère donc important de vérifier l'impact sur ce qui sera défini comme étant le comportement séquentiel des répondants. Finalement, un changement associé à l'option de non-réponse sera réalisé et fera l'objet d'une analyse approfondie.

### **Format des questionnaires de l'EEE et qualités psychométriques**

Depuis quelques années, l'étude des qualités psychométriques des instruments associés à l'EEE suscite l'intérêt de différents acteurs en enseignement universitaire. L'EEE s'insère dans une démarche générale visant à assurer la qualité en enseignement supérieur basée sur ce qu'en pensent ceux qui en bénéficient. L'EEE est utilisée à des fins pédagogiques, mais également à des fins d'évaluation des ressources professorales (Donovan, Mader & Shinsky, 2006; Fontaine, 2009). Les évaluations sont notamment considérées lors de l'embauche ou de l'obtention de promotion et ont parfois des impacts sur l'obtention de contrats d'enseignement. Elles

sont aussi un objet de recherche en enseignement supérieur et suscitent l'intérêt des chercheurs. Finalement, la qualité de l'enseignement devient maintenant un critère pour la sélection des cours et des universités par les parents et les étudiants. Ces différentes décisions prises par les acteurs en enseignement universitaire nécessitent que les instruments utilisés pour l'EEE soient valides et fiables, car ce sont souvent à propos de ces aspects de la mesure que se font les critiques les plus virulentes (Detroz, 2008).

L'étude des qualités psychométriques des instruments associés à l'EEE a ses particularités. Premièrement, l'EEE est un exercice récurrent. Par exemple, dans l'université à l'étude, chacune des activités d'enseignement est évaluée à partir d'un instrument composé de plusieurs énoncés; dans le cas des cours en classe ordinaire, le questionnaire comprend 26 énoncés<sup>1</sup> et, si les étudiants sont inscrits à temps plein (5 cours), la démarche totalise 130 énoncés (5 x 26) chaque trimestre, sans compter les commentaires écrits, le cas échéant. La démarche devient alors fastidieuse et une diminution de l'intérêt à répondre peut s'installer chez les répondants (Dolmans & Ginns, 2005). En ce sens, l'EEE diffère des épreuves, qui sont effectuées une seule fois auprès d'une même personne.

Deuxièmement, l'étude des qualités psychométriques s'insère dans un contexte institutionnel réel, ce qui impose certaines contraintes aux dispositifs de validation qui peuvent être mis en place. Les choix effectués doivent être optimaux en matière de temps et de mobilisation de ressources, et doivent permettre de poursuivre les opérations normales de l'institution. Les données recueillies possèdent une forte validité externe en ce qu'elles sont issues de situations réelles d'enseignement; en contrepartie, ce sont des données secondaires qui ne se prêtent pas à la manipulation expérimentale. Ces contraintes justifient les choix qui ont été réalisés dans le cadre de la présente recherche.

Troisièmement, l'étude des qualités psychométriques s'insère dans un contexte où les instruments utilisés pour l'EEE possèdent des formats et des modes de passation variés. Or, les choix effectués dans les différentes universités peuvent avoir un impact sur les qualités psychométriques des instruments. Une certaine attention au format et aux modes de passation est donc nécessaire. Ainsi, à l'université ciblée, le mode de passation du questionnaire utilisé pour l'EEE a été modifié d'une modalité d'administration papier à une passation en ligne et l'impact de ce changement est documenté (Harvey & Hébert, 2012). Globalement, des taux supérieurs à

87% sont rapportés pour les modalités de passation en classe ordinaire et virtuelle. Cependant, avec l'implantation de l'évaluation en ligne, les taux de réponse sont passés de 84,9% à 58,6%; cette diminution des taux de réponse lorsque la passation se déroule en ligne plutôt qu'en classe ordinaire est d'ailleurs largement documentée (Donovan et al., 2006).

Par ailleurs, l'un des aspects les plus importants à considérer lors du processus de conception d'une version abrégée est la réduction du nombre de questions; celle-ci peut avoir un impact direct sur la dimensionnalité de l'instrument, sur sa fiabilité, sur l'appréciation de la qualité au moyen de commentaires écrits ou sur le comportement séquentiel des répondants.

### ***La dimensionnalité et la fiabilité des questionnaires de l'EEE***

Généralement, la documentation (p. ex., Marsh & Bailey, 1993) suggère que les instruments utilisés pour l'EEE sont valides et fiables. Il existe cependant un problème connu lié à la dimensionnalité de ces questionnaires (d'Apollonia & Abrami, 1997). Ainsi, les outils conçus sont constitués de plusieurs dimensions souvent très corrélées entre elles, si bien que l'étude de la dimensionnalité ne relève la présence que d'une dimension dominante, qui, à l'occasion, peut être accompagnée de dimensions de moindre importance. À titre d'exemple (Harvey & Hébert, 2012), le questionnaire utilisé à l'université sous considération est constitué de six dimensions. Cependant, les évaluations réalisées par les étudiants ne permettent pas de porter un jugement valide et fiable sur chacune de ces dimensions prises indépendamment. Une analyse factorielle ne révèle que la présence de deux facteurs orthogonaux: un facteur principal, qui explique à lui seul 61% de la variance totale, est accompagné d'un second facteur, qui n'explique quant à lui que 5% de la variance. Par ailleurs, seul le facteur principal possède la fiabilité adéquate pour qu'une décision puisse être basée sur son évaluation. Ce facteur est généralement interprété comme étant lié à la mesure d'une compétence générale en enseignement ou bien au professionnalisme de la ressource professorale. Cette structure dimensionnelle est indépendante de la modalité de passation du questionnaire, que cette modalité soit réalisée en ligne ou en classe à partir d'une version papier du questionnaire (Harvey & Hébert, 2012). La fiabilité de l'ensemble des items est de 0,96, basée sur l'alpha de Cronbach.

Est-il alors nécessaire de maintenir un questionnaire composé de 21 items et de 6 dimensions si l'ensemble ne permet qu'un jugement global sur une dimension unique? Étant donné que l'exercice est récurrent et fasti-

dieux, est-il possible de concevoir un questionnaire abrégé qui réalisera la même fonction, et ce, sans altérer les qualités psychométriques de l'instrument? Notre hypothèse principale ( $H_1$ ) est qu'une version abrégée allègera le processus d'évaluation, le rendant moins fastidieux, et permettra d'augmenter les taux de réponse sans altérer les qualités psychométriques de l'instrument.

Notre hypothèse  $H_{1a}$  est qu'un facteur principal fort structurera les réponses aux deux versions. Ce facteur devrait expliquer plus de 60% de la variance totale (Harvey & Hébert, 2012). Tous les énoncés devraient posséder une saturation forte avec celui-ci. Une forte fiabilité du construit mesuré est attendue dans les deux cas. La présence d'un second facteur n'est pas anticipée.

Le concept de validité (Messick, 1995) ne se limite pas à l'étude de la dimensionnalité ni de la fiabilité d'un outil. Il s'applique plus largement à toutes inférences basées sur des observations ou des comportements cohérents et systématiques associés au processus de réponse. Dans le présent article, une attention particulière sera portée à l'appréciation de la qualité au moyen de commentaires écrits, au comportement séquentiel et à l'option de non-réponse, qui pourraient être modifiés à la suite de la production d'une version abrégée.

### ***L'appréciation de la qualité au moyen de commentaires écrits***

La version abrégée, en réduisant le nombre de questions sur un thème donné, peut modifier le rappel d'information pertinente; il peut s'ensuivre une diminution de l'appréciation au moyen de commentaires écrits, une augmentation du nombre d'énoncés sans réponse et, par conséquent, un changement global dans le comportement séquentiel des répondants.

Tel que nous l'avons mentionné précédemment, le format de l'instrument utilisé à l'université ciblée est hybride; il se compose de questions fermées qui s'accompagnent d'une section ouverte permettant aux étudiants d'exprimer dans leurs propres mots leur appréciation du cours qu'ils ont suivi.

Dans la mesure où les commentaires donnent accès aux représentations des étudiants et que ceux-ci se révèlent particulièrement utiles aux professeurs pour l'amélioration de leurs cours, cet aspect sera documenté dans le présent article; l'impact potentiel sur la qualité du cours représente une conséquence directe du changement (validité de conséquence).

En particulier, avec la réduction du nombre de questions de l'outil, les commentaires écrits prennent une signification accrue. Ils précisent l'interprétation à donner à un choix de réponse. Il est difficile de prédire a priori quel sera l'impact d'un questionnaire abrégé sur l'énonciation des commentaires par les étudiants.

Deux hypothèses sont examinées dans la présente recherche. L'hypothèse  $H_2$  est que l'étudiant qui désire exprimer une opinion à propos de l'appréciation d'un cours l'énoncera, et ce, peu importe le format que prendra le questionnaire. Sous cette hypothèse, les deux questionnaires susciteront une quantité équivalente de commentaires.

### ***Le comportement séquentiel***

À l'opposé, l'hypothèse  $H_3$  stipule que le format du questionnaire joue un important rôle d'élicitation en raison des questions posées. Ainsi, dans certains questionnaires, l'organisation séquentielle des items est telle qu'elle favorise l'activation en mémoire de thèmes et favorise le rappel à des questions subséquentes. Un facteur de mémorisation est donc en jeu. Une organisation opposée serait par exemple la présentation au hasard des énoncés. À notre connaissance, la plupart des questionnaires d'EEE possèdent une telle structure séquentielle des énoncés. Les énoncés concernant par exemple l'évaluation des travaux et des examens sont consécutifs, tout comme ceux concernant le respect du plan de cours et l'organisation du cours. Cette organisation favorise le rappel de l'information pertinente et minimise ainsi le nombre d'énoncés sans réponse (Carley-Baxter, Peytchev & Black, 2010). L'impact sur les questions sans réponse sera discuté dans une section ultérieure.

Pour l'heure, considérant que le questionnaire abrégé contient moins de questions, celui-ci suscitera peut-être moins de commentaires que le questionnaire original. Afin de minimiser cet impact d'un nombre restreint de questions, le questionnaire abrégé a été conçu afin de solliciter une énonciation de commentaires par des mots clés ravivant en mémoire les aspects de l'enseignement visés par la question; la fonction d'élicitation est donc maintenue au moyen de ces mots clés.

L'étude du comportement séquentiel permettra d'effectuer cette étude de la modification de l'organisation séquentielle des questions en fonction des versions originale et abrégée du questionnaire. Le comportement séquentiel est défini comme étant la propension d'un étudiant à répondre

la même réponse à un énoncé ainsi qu'à l'énoncé qui le précède. Ce comportement produit alors des séquences identiques de réponses consécutives. Deux facteurs distincts peuvent expliquer un tel comportement séquentiel. Le premier facteur atteste que l'étudiant n'a pas changé d'opinion d'une question à l'autre. En ce sens, le comportement séquentiel est lié à la structure factorielle du questionnaire. Le second facteur est lié à une certaine mémorisation de la réponse précédente, qui a alors un impact sur la réponse suivante ; l'étudiant produit ainsi une réponse identique. La question qui se pose alors est de savoir s'il y aura des changements importants dans le comportement séquentiel des répondants en fonction de la version (abrégée ou originale) du questionnaire.

Deux méthodes sont proposées pour mettre en évidence les changements potentiels de cette organisation séquentielle : l'analyse séquentielle et l'analyse factorielle exploratoire. L'analyse séquentielle a pour but de rechercher les séquences de réponses consécutives et de les comptabiliser. Elle permet de mieux comprendre le contexte dans lequel les choix de réponse sont utilisés. Par exemple, l'option de non-réponse est-elle plutôt précédée d'une réponse « tout à fait en accord » ou d'une réponse « tout à fait en désaccord » ? Quel est le type de réponse qui lui succède ?

La seconde méthode repose sur un modèle d'analyse factorielle confirmatoire. Green et Hershberger (2000, équation 9) ont proposé un modèle structurel autorégressif où chaque item est modélisé à partir d'un facteur principal, auquel s'additionne la réponse pondérée à l'item précédent. Le pondérateur sert alors d'estimation de l'importance du facteur autorégressif (facteur de mémorisation).

### ***L'option de non-réponse***

L'option de non-réponse est un choix qui est offert aux répondants et qui se présente sous une forme telle que « ne s'applique pas », « je ne sais pas », « je préfère ne pas répondre » ou autre. Une telle option est utilisée lorsqu'il apparaît improbable que tous les répondants répondent à tous les items (Holman, Glas, Lindeboom, Zwinderman & de Haan, 2004). Selon Feick (1989), l'origine de ce type de réponse est attribuable à la fois à des caractéristiques des répondants et des items. Ainsi, ce choix de réponse se produit lorsque l'item apparaît non pertinent par rapport au construit mesuré, ambigu ou mal énoncé. Le répondant est alors ambivalent dans son choix de réponse et n'arrive pas à se faire une opinion sur la question posée. Dans le cas de l'EEE, il se trouve alors dans une zone grise entre la

satisfaction et l'insatisfaction par rapport à la question posée. Bernard (2011) suggère qu'un questionnaire de l'EEE doit inclure une telle option au cas où un énoncé ne convient pas à la situation spécifique du cours. Ces réponses sont alors traitées comme des données manquantes et certaines stratégies pour les analyser sont proposées (Holman et al., 2004; Hwang, Petrolia & Interis, 2014).

Certains auteurs (Hwang et al., 2014) considèrent également qu'il existe une relation entre ce choix de réponse et la conséquentialité du questionnaire. La conséquentialité est une mesure du degré par lequel un questionnaire aura un impact ultérieur sur les politiques. Les répondants qui perçoivent les conséquences d'un questionnaire ont tendance à moins utiliser l'option de non-réponse. Dans le cas de l'EEE, les conséquences sont généralement connues; le questionnaire a une fonction de rétroaction auprès des ressources professorales et constitue une mesure incitative à l'amélioration continue de l'enseignement. Il peut également avoir des conséquences sur la carrière professorale et sur l'obtention de tâches d'enseignement. Les modifications proposées dans la présente version abrégée n'ont cependant pas d'impact sur la conséquentialité du questionnaire.

Au moins deux types d'énonciations ont été répertoriés. La première énonciation, «ne s'applique pas» (*not applicable = n/a*), suggère que l'item est inadéquat dans la situation. La seconde, «je ne sais pas» (*do not know/no opinion/not sure*), exprime plutôt l'ambivalence du répondant. Afin de concilier ces deux explications, la version abrégée offre un choix de réponse qui combine ces deux aspects, bien que ceux-ci soient différents. Ce choix est «ne s'applique pas/je ne sais pas».

Cet ajout rend plus cohérent le choix de ce type de réponse. A priori, il est difficile de prédire si cette modification engendrera une augmentation ou une diminution de ce type de réponse. Si ce choix est plus cohérent avec les énoncés et l'opinion ambivalente des répondants, il est plausible que sa fréquence augmente. Krosnick et ses collaborateurs (2002) précisent que ce choix de réponse permet à ceux qui n'ont pas d'opinion de ne pas être obligés de répondre au hasard parmi les choix proposés; dans un tel cas, ce choix réduit la variance aléatoire dans les réponses, et améliore la validité et la fiabilité du sondage. À l'opposé, il peut représenter une invitation à ne pas communiquer sa véritable opinion pour ceux qui en ont une; cela réduit alors la qualité du sondage, donc sa validité et sa fiabilité.

De plus, la position séquentielle de l'option de non-réponse est modifiée dans le présent questionnaire; alors qu'elle faisait suite à l'option «tout à fait en désaccord» dans le questionnaire original, elle précède maintenant l'option «tout à fait d'accord» (voir Tableau 1). La distance proximale étant considérée, l'option de non-réponse, qui était associée à l'option «tout à fait en désaccord» dans le questionnaire original, est maintenant associée à l'option «tout à fait d'accord» dans la version abrégée.

Tableau 1  
*Exemple d'une question et des choix de réponse  
des versions originale et abrégée*

Version originale	Version abrégée
Le cours est bien structuré.	Le cours est bien organisé (clarté du plan de cours, préparation, contenu et structure du cours, charge de travail requise).
0. Ne s'applique pas	○ Ne s'applique pas/Je ne sais pas
1. Tout à fait en désaccord	○ Tout à fait d'accord
2. Plutôt en désaccord	○ Plutôt d'accord
3. Plutôt en accord	○ Plutôt en désaccord
4. Tout à fait en accord	○ Tout à fait en désaccord
Commentaires <sup>1</sup> :	
<hr/>	
<hr/>	
<hr/>	

1. La zone «Commentaires» est identique pour les deux versions.

Étant donné que la démarche abrégée modifie ce choix de réponse, le présent article documentera le contexte d'occurrence de ce choix de réponse pour les versions originale et abrégée. Une attention spéciale sera accordée à la prévalence de ce choix de réponse en fonction des questions. Ainsi, ce choix de réponse est-il associé à certaines questions en particulier, ce qui serait alors une indication que celles-ci sont inadéquates ou ambiguës dans la situation?

### ***La validité et la fiabilité des décisions***

À ce stade-ci, il est impératif d'effectuer un rappel des notions de validité et de fiabilité afin de préciser les qualités psychométriques recherchées dans le contexte de l'étude des questionnaires d'EEE, et ce, pour expliciter les choix méthodologiques ultérieurs.

La validité est considérée comme un jugement global porté sur l'ensemble des preuves empiriques et théoriques qui atteste que les interprétations et les actions basées sur les résultats à un test sont adéquates (American Educational Research Association, 1999; Messick, 1995). Dans cette définition, la validité est un concept unifié intégrant des aspects liés au contenu et à la représentativité des évaluations, à la structure et au construit, à des critères internes et externes, ainsi qu'à la généralisabilité et aux conséquences de l'évaluation. Cette conception de la validité concilie l'apport de considérations théoriques et empiriques, ainsi que les conséquences sociales et éthiques de l'utilisation de l'outil sous validation. Cette définition inclut les différentes formes de validité, dont la validité de contenu, de construit, de critère et de conséquence.

La validité de contenu est basée sur l'évaluation des items par des experts qui en évaluent la pertinence et la représentativité. La validité de construit réfère à l'adéquation entre la structuration attendue des évaluations et celle empiriquement obtenue. La validation à partir de critères externes est également utilisée. Les critères externes de référence sont variés. Il peut s'agir d'une évaluation concomitante d'une prestation universitaire réalisée par exemple par un ou des observateurs *in situ*, par un portfolio d'enseignement, par un enseignement antérieur, par l'utilisation de méthodes ou de formulaires différents (Detroz, 2008) et par l'utilisation d'un groupe de discussion (*focus group*) ou de la tenue de journaux réflexifs (Huxham, Laybourn, Cairncross, Gray, Brown, Goldfinch & Earl, 2008). La validité de conséquence porte sur les conséquences sociales et éthiques potentielles associées à l'utilisation du test, à son interprétation et aux actions subséquentes qu'il engendre.

D'autres définitions de la validité ont été récemment proposées (voir André, Loye & Laurencelle, 2015 pour une recension). Par exemple, Cizek (2012) argumente que la validation d'un instrument de mesure est une action qui doit être dissociée de la justification sociale et éthique de son utilisation et que, par conséquent, ces aspects ne doivent pas faire partie de la définition du concept de validité.

Borsboom, Mellenbergh et van Heerden (2004) considèrent que les questions liées à la validité sont d'ordre ontologique, et qu'elles sont associées à l'existence même d'un attribut et de son lien de causalité avec les résultats des évaluations. Ils considèrent qu'un instrument procure une mesure valide d'un attribut lorsqu'il est possible de démontrer que cet

attribut existe et que des fluctuations de celui-ci provoquent (dans le sens de causalité) des changements dans les indicateurs utilisés. Cependant, cette conception de la validité est difficilement applicable à l'analyse des évaluations réalisées dans le cadre de l'EEE. Ainsi, les EEE ont une existence qui est liée à l'action de suivre un cours ; elles sont ainsi des données invoquées. Elles ne peuvent, rétroactivement, faire l'objet de manipulations afin de provoquer des changements et d'observer des relations causales.

Par conséquent, la plupart des méthodes de validation répertoriées utilisent des procédés basés sur l'inspection de la distribution des statistiques descriptives et sur la structuration des résultats à partir d'analyses factorielles exploratoire et confirmatoire afin de déterminer et de mettre en relation les facteurs mesurés durant l'enseignement. En ce sens, la plupart des études recensées s'insèrent dans une conception de la validité cohérente avec la définition de Messick (1995).

La fiabilité réfère à la précision et à la consistance interne d'un instrument, c'est-à-dire la proportion de la variance attribuable au score vrai par rapport au total de la variance (Brown, 2006 ; Raykov, 2004). Les méthodes disponibles pour établir la fiabilité des évaluations sont diverses (Redding, Maddock & Rossi, 2006 ; Saupe & Eimers, 2012), mais ne sont pas toutes applicables à l'analyse d'EEE. Certaines requièrent l'utilisation de deux mesures, telles que l'utilisation de formes parallèles et équivalentes d'un même test (p. ex, version A par rapport à version B) ; cette condition ne s'est pas avérée possible dans un contexte institutionnel réel, sujet à des limites imposées par les dispositifs technologiques en place.

La méthode du test-retest nécessite également que deux mesures identiques soient réalisées avec le même instrument ; cette méthode ne s'applique à l'EEE que si l'on considère que les résultats d'un enseignement réalisé lors d'un trimestre sont comparables à ceux d'un second trimestre. La corrélation entre deux enseignements successifs peut alors être utilisée. Cependant, dans le cadre de l'EEE, cette méthode a ses limites, car il n'y a jamais deux enseignements identiques. Cette méthode n'a pas été retenue.

D'autres méthodes existent et ne nécessitent qu'une prise de mesure. Elles sont basées sur le calcul de la consistance interne (p. ex., le coefficient alpha de Cronbach), sur la fiabilité fractionnée de l'échelle (corrélation d'une moitié de l'échelle avec l'autre) ou sur l'analyse de la variance

(p. ex., étude de la généralisabilité). La méthode basée sur le calcul de la consistance, en raison de sa simplicité et du fait qu'elle permet de bien atteindre les objectifs de la présente recherche, a été retenue.

### ***La conception d'une version abrégée***

Certaines universités (Laurentienne, McGill et Yale Law) ont déjà adopté des versions de l'EEE avec un nombre réduit d'items se composant parfois de questions fermées ou ouvertes. Dans la présente recherche, la conception d'une version abrégée a été réalisée avec le souci de maintenir les qualités psychométriques du questionnaire original. Cette condition n'est pas triviale. Ainsi, l'absence de cette directive peut donner naissance à un questionnaire abrégé fort différent de l'original. Par exemple, Huxham et ses collaborateurs (2008) montrent qu'un questionnaire constitué uniquement de 25 questions fermées n'est pas équivalent à un formulaire à rétroaction rapide constitué de 3 questions ouvertes. Gal et Gal (2015) suggèrent également qu'une version B d'un questionnaire A peut être très différente si cette version utilise des items, des thèmes et un vocabulaire différents; la démonstration a été faite à partir d'une version B ayant été générée par des étudiants utilisant leur propre vocabulaire et leur propre conception de l'évaluation d'un cours. Par conséquent, le maintien d'un vocabulaire et d'items similaires est recommandé pour éviter la perte des qualités psychométriques dans le processus.

Une des stratégies possibles pour la conception d'une version abrégée d'un instrument consiste à sélectionner un sous-ensemble pertinent d'énoncés à partir de l'ensemble original. Cette sélection peut être basée sur les pondérateurs d'une analyse factorielle ou sur les indices de difficulté et de discrimination issus, par exemple, de la théorie de réponse aux items. Cependant, dans le cas présent, cette stratégie risque de générer des sous-ensembles d'items qui, globalement, sous-représentent le construit mesuré.

Une deuxième stratégie, qui consiste à adopter quelques énoncés généraux auxquels s'ajoute un nombre plus ou moins grand d'énoncés choisis par la ressource professorale, a aussi été rejetée; elle risque également de générer des sous-ensembles peu représentatifs du construit. De plus, la gestion des multiples versions ainsi produites apparaît difficilement optimale en ce qui concerne les coûts et les bénéfices.

Finalement, Dolmans et Ginns (2005) ont produit un questionnaire court à cinq dimensions en définissant deux ou trois questions par dimension. Cependant, comme nous en avons discuté antérieurement, cette stratégie tend à générer des facteurs intercorrélés.

Par conséquent, ces stratégies ont été écartées au profit d'une stratégie où seuls des énoncés généraux mesurant un facteur unique sont produits. Chaque énoncé effectue la synthèse des énoncés associés à l'une des dimensions du questionnaire original. Chaque énoncé s'accompagne d'un espace réservé pour la formulation de commentaires portant sur la question générale. Le tableau 1 présente un exemple d'énoncé de ce questionnaire abrégé. Dans la version originale, la dimension *Organisation et clarté* du cours regroupe cinq énoncés. Ceux-ci portent sur 1) la clarté du plan de cours, 2) son respect, 3) la structure du cours, 4) le respect de la description du cours, et 5) la pertinence des moyens pédagogiques utilisés. L'énoncé général présenté au tableau 1 effectue une synthèse de ces cinq énoncés : *Le cours est bien organisé*. Entre parenthèses sont précisés les aspects spécifiques visés : la clarté du plan de cours, la préparation, le contenu et la structure du cours, ainsi que la charge de travail requise. L'étudiant doit indiquer, sur une échelle de Likert à cinq modalités, sa position, qui se situe entre être « tout à fait d'accord » et « tout à fait en désaccord » avec l'énoncé. Il a également la possibilité de répondre « ne s'applique pas/je ne sais pas » à l'énoncé. Une zone dédiée à la formulation d'un commentaire permet à l'étudiant de préciser sa pensée sur la question générale. Cette combinaison d'une question générale fermée et d'un commentaire ouvert assure que le questionnaire est représentatif du construit original. En tout, le questionnaire comprend 10 questions, dont 8 portent sur la prestation d'enseignement et 2 sur le contexte matériel (locaux, etc.) et la pertinence du cours dans le programme. Seuls les 8 premiers énoncés sont utilisés pour prendre une décision sur la prestation de la ressource professorale.

Comme nous en avons discuté auparavant, le choix de réponse « ne s'applique pas » a été modifié par « ne s'applique pas/je ne sais pas ». Ce choix figure également au début de l'échelle de Likert et aucune valeur n'y est associée. Une autre modification à l'échelle est qu'il n'y a plus de valeurs numériques associées aux choix, alors qu'elles apparaissaient en ordre croissant dans l'échelle originale. De plus, la position séquentielle

de l'option de non-réponse est modifiée ; alors qu'elle faisait suite à l'option « tout à fait en désaccord » dans le questionnaire original, elle précède maintenant l'option « tout à fait d'accord ».

Il est possible que ces modifications aient un impact sur la manière dont les étudiants répondent aux énoncés, en particulier sur le comportement séquentiel des répondants. Une analyse des taux de satisfaction et une analyse de la fréquence des choix de non-réponse seront réalisées en parallèle avec une analyse du comportement séquentiel.

## Méthode

### *Les questionnaires*

Deux questionnaires sont utilisés pour cette étude : le questionnaire original et le questionnaire abrégé. Le questionnaire original comprend six dimensions qui sont évaluées par les étudiants. Celui-ci est constitué au départ de 26 items. Il se décline comme suit (voir Tableau 3) : *Contexte du cours* (items 1 à 5), *Organisation et clarté* (items 6 à 13), *Dynamisme, intérêt et habiletés d'enseignement* (items 14 à 17), *Interaction avec les étudiants et étudiantes* (items 18 à 20), *Évaluation et rétroaction* (items 21 à 24) et *Appréciation générale* (items 25 et 26). L'étudiant sondé doit marquer son degré d'accord ou de désaccord aux items au moyen d'une échelle de Likert à cinq modalités (voir Tableau 1).

Il faut préciser que les items de la dimension *Contexte du cours* sont indépendants de la ressource professorale. Des items de ce type sont aussi présents dans d'autres instruments (Fontaine, 2009 ; Thorpe, 2002). À l'université ciblée, ils balisent la décision prise par l'institution au sujet de la ressource professorale, mais ne sont pas considérés dans le volet sommatif du test.

Le questionnaire abrégé se compose d'une dimension composée de 8 items, à laquelle s'ajoutent des questions sur le contexte et sur la pertinence du cours dans le programme. L'étudiant doit marquer son degré d'accord ou de désaccord aux items au moyen d'une échelle de Likert à cinq modalités (voir Tableau 1).

### ***Les échantillons***

Deux échantillons de données ont été constitués. Le premier, constitué au trimestre d'hiver 2014 à partir de la version originale du questionnaire, est composé de 6870 répondants répartis dans 517 classes. Les classes où deux professeurs ont enseigné durant le trimestre sont comptabilisées comme des classes différentes, car des évaluations différenciées ont eu lieu. Le second échantillon, constitué au trimestre d'hiver 2015 au moyen de la version abrégée, est composé de 7257 répondants répartis dans 568 classes. Les données proviennent de huit départements et de deux campus de l'Université du Québec à Rimouski, situés à Lévis et à Rimouski.

### ***La procédure d'administration des instruments***

Au maximum 24 jours précédant les cinq derniers jours ouvrables du trimestre en cours, les étudiants peuvent remplir le questionnaire en ligne. Le professeur responsable d'une activité d'enseignement explique la procédure à suivre aux étudiants avant le début de la période d'évaluation. Des rappels sont respectivement envoyés aux ressources et aux étudiants lorsque la moitié et les trois quarts de la période d'évaluation sont écoulés. Au début du trimestre suivant, les résultats de l'évaluation sont transmis au professeur, puis analysés à des fins pédagogiques par le comité de gestion du programme auquel l'activité d'enseignement est rattachée. Les résultats sont également considérés quant à certains aspects liés à la carrière professorale (engagement, promotion et formation). Ils détaillent la fréquence de réponse pour chacun des items et, le cas échéant, pour chacune des cinq dimensions. Un résultat global peut être calculé.

### ***La description des étapes de validation et des analyses***

L'ensemble des opérations de conception et de validation du questionnaire abrégé est supervisé par un comité institutionnel dont le mandat est de revoir les pratiques, la réglementation ainsi que le questionnaire relatif à l'évaluation de l'enseignement. La présente étude s'insère en marge des travaux de ce comité.

#### ***La passation de la version abrégée (hiver 2015) et de la version originale (hiver 2014)***

La version abrégée a été utilisée à l'hiver 2015 pour l'évaluation de toutes les activités d'enseignement offertes en classe ordinaire aux campus de l'université ciblée situés à Rimouski et à Lévis. La version originale a été

utilisée lors du trimestre d'hiver 2014. Les deux questionnaires étaient disponibles uniquement en version en ligne. Dans les deux cas, les étudiants ont été invités à évaluer chacun des cours auxquels ils étaient inscrits. Une invitation à procéder à l'évaluation figurait sur la plateforme informatisée d'apprentissage personnel des étudiants (Moodle). Cette invitation demeurait tant que l'étudiant n'avait pas procédé à l'évaluation.

### ***La validation du contenu***

La validation du contenu des items consiste à vérifier auprès des experts que l'ensemble des énoncés est pertinent, bien formulé et représentatif du contenu. La version abrégée a fait l'objet d'une telle validation auprès d'experts internes et externes à l'institution ainsi qu'auprès de membres de la communauté universitaire qui ont accepté de se prêter à l'exercice. Tout autre commentaire au sujet de l'échelle ou du questionnaire a également été recueilli lors des séances de consultation. Le questionnaire abrégé dans sa forme actuelle est le résultat de ce processus de conception et de validation.

### ***Les taux de satisfaction et les taux de réponse***

Les taux de satisfaction et les taux de réponse quant à l'utilisation de l'une et l'autre version du questionnaire sont compilés. Comme nous l'avons mentionné précédemment, une augmentation des taux de réponse est anticipée avec la version abrégée. En contrepartie, aucune fluctuation des taux de satisfaction n'est anticipée entre les deux versions ( $H_1$ ).

### ***La validation du construit et la fiabilité***

Afin de répondre à l'hypothèse  $H_{1a}$ , une analyse factorielle est effectuée pour chacune des versions du questionnaire. Seuls les énoncés considérés dans l'évaluation de la ressource professorale sont utilisés ; les énoncés liés au contexte sont exclus. L'alpha de Cronbach est utilisé comme indicateur de la fiabilité.

### ***L'analyse des commentaires ( $H_2$ et $H_3$ )***

L'analyse des commentaires comporte deux volets quantitatifs. Le premier volet s'intéresse au nombre de commentaires par répondant et par question. Le second volet concerne la longueur des commentaires par répondant et par question.

### *Le comportement séquentiel*

#### *Analyse séquentielle*

À ce stade-ci, seules des séquences d'ordre 1 sont recherchées. Une séquence d'ordre 1 ne considère que les chaînes à deux éléments. Ainsi, la probabilité conditionnelle d'observer un choix de réponse en tenant compte du choix précédent est calculée pour chacun des choix de réponse. Étant donné qu'il y a 5 choix de réponse (0 à 4), il y a 25 séquences possibles présentées dans l'ensemble suivant [(0,0) (0,1) (0,2) (0,3)... (4,4)]. Le couple (4,4) signifie que l'étudiant a répondu 4 (« tout à fait en accord ») à deux questions consécutives. L'analyse séquentielle compile toutes les séquences possibles de réponses et calcule la probabilité d'occurrence de chacune. Une matrice de transition d'un choix de réponse à l'autre permet de synthétiser l'information.

#### *Analyse factorielle confirmatoire*

L'analyse séquentielle confond la présence du facteur principal et du facteur de mémorisation dans la production d'une réponse identique, donc elle ne permet pas d'estimer l'importance du facteur de mémorisation. Par conséquent, dans un second temps, une analyse factorielle confirmatoire est réalisée. Un modèle ayant un facteur principal et une composante auto-régressive est élaboré. Celui-ci s'inspire de Green et Hershberger (2000, équation 9).

Par exemple, pour les items 1 à 3 du questionnaire, les équations structurelles se présentent ainsi :

$$Q1 = a_1 F1 + e_1 ;$$

$$Q2 = a_2 F1 + \text{bêta} Q1 + e_2 ;$$

$$Q3 = a_3 F1 + \text{bêta} Q2 + e_3 .$$

Ainsi, la réponse donnée à l'item Q2 est une fonction du facteur général F1 pondérée par  $a_2$  du choix effectué à la réponse Q1 et de l'erreur  $e_2$ . Le paramètre bêta est le facteur de mémorisation qui détermine l'importance de la réponse Q1 pour Q2. Ainsi, compte tenu de la contribution de la structure factorielle du questionnaire, quelle est l'importance du facteur de mémorisation (bêta) dans les réponses observées? Notons qu'un même paramètre de mémorisation est utilisé pour tous les items. Cette stratégie permet de constituer un modèle simple et de démontrer la présence simultanée du facteur principal et du facteur de mémorisation, puis d'en estimer les paramètres respectifs.

### *L'analyse de l'option de non-réponse*

La fréquence de ce type de réponse est calculée globalement et en fonction de chaque question. Les statistiques des deux versions sont comparées.

## Résultats

### *Les taux de réponse et les taux de satisfaction*

Le taux de réponse du questionnaire original à l'hiver 2014 est de 48,9% (voir Tableau 2). En 2010-2011, le taux de réponse observé était de 58,6% (Harvey & Hébert, 2012). Une diminution du taux de réponse est donc observée depuis cette période.

Tableau 2  
*Taux de réponse*

Version du questionnaire	Répondants potentiels	Répondants réels	Taux de réponse (%)
Originale	14 064	6 870	48,9
Abrégée	14 481	7 257	50,1

Quant au taux de satisfaction de la version originale, il est de 3,46/4 (86,6%) (excluant la dimension *Contexte du cours*), avec un écart-type de 0,62 (voir Tableau 3). Les taux de satisfaction enregistrés pour chacune des dimensions et globalement sont similaires à ceux qui sont documentés pour cette version du questionnaire (Harvey & Hébert, 2012).

Tableau 3  
*Taux de satisfaction – Questionnaire original*

Dimension	Hiver 2014		
	N <sup>bre</sup> d'items	Moyenne	Écart-type
Contexte du cours	5	3,60	0,45
Organisation et clarté	8	3,51	0,59
Dynamisme, intérêt et habiletés	4	3,45	0,70
Interaction avec les étudiants et étudiantes	3	3,45	0,73
Évaluation et rétroaction	4	3,42	0,66
Appréciation générale	2	3,41	0,80
<i>Globale</i>	<i>21</i>	<i>3,46</i>	<i>0,62</i>

Le taux de réponse du questionnaire abrégé est de 50,1% (voir Tableau 2). Le taux de satisfaction (voir Tableau 4) est de 3,49/4 (87,2%). En moyenne, tous les aspects des cours sont appréciés, avec un taux de satisfaction se situant entre 3,35/4 (83,7%) et 3,68/4 (92,0%). Ces taux sont comparables à ceux obtenus avec la version originale. L'augmentation du taux de réponse stipulé en  $H_1$  ne s'est donc pas produite.

Tableau 4  
*Taux de satisfaction – Questionnaire abrégé*

Question	Hiver 2015	
	Moyenne	Écart-type
Q1. Le cours est bien organisé.	3,43	0,78
Q2. Le plan de cours a été respecté.	3,67	0,62
Q3. L'enseignant maîtrise la matière.	3,68	0,61
Q4. L'enseignant démontre de bonnes habiletés d'enseignement.	3,40	0,84
Q5. L'enseignant évalue avec rigueur les apprentissages.	3,43	0,81
Q6. L'enseignant est disponible en dehors des heures de cours.	3,62	0,64
Q7. Le cours permet de réaliser des apprentissages significatifs.	3,43	0,79
Q8. Dans l'ensemble, je suis satisfait de ce cours.	3,35	0,85
Q9. Les conditions matérielles et temporelles du cours sont adéquates.	3,53	0,70
Q10. Le cours est pertinent dans le programme.	3,55	0,73
<i>Globale</i>	<i>3,49</i>	<i>0,60</i>

### ***La validation du construit et la fiabilité***

Un seul facteur émerge de l'analyse factorielle réalisée à partir de la version originale. La valeur de la racine latente associée au premier facteur est de 13,3 (sur 21) et explique 63% de la variance totale. La fiabilité de l'ensemble est de 0,96 (alpha de Cronbach). Ces résultats sont comparables à ce qui a été documenté antérieurement.

Le tableau 5 présente les pondérateurs associés au facteur observé. Les pondérateurs sont forts et fluctuent entre 0,61 et 0,88. Tous les énoncés sont donc associés à ce facteur unique. Comme nous l'avons mentionné antérieurement, cette structure valide les décisions basées sur le résultat global au questionnaire; elle ne permet pas de porter un jugement sur l'une ou l'autre des sous-dimensions prises d'une manière indépendante.

Tableau 5

*Analyses factorielles exploratoire et confirmatoire (avec un facteur principal et un facteur autorégressif de mémorisation – coefficients standardisés)*

Questionnaire original <sup>1</sup>					
N° d'énoncé	Analyse exploratoire	Analyse confirmatoire			
		Facteur principal	Facteur de mémorisation	Erreur	R <sup>2</sup>
6	0,72	0,69	s. o.	0,72	0,48
7	0,72	0,58	0,17	0,71	0,50
8	0,84	0,73	0,14	0,56	0,69
9	0,81	0,63	0,20	0,61	0,63
10	0,83	0,73	0,14	0,56	0,68
11	0,64	0,45	0,17	0,80	0,36
12	0,75	0,56	0,17	0,74	0,46
13	0,83	0,68	0,17	0,59	0,66
14	0,77	0,58	0,20	0,67	0,56
15	0,85	0,76	0,11	0,53	0,76
16	0,87	0,71	0,18	0,50	0,75
17	0,81	0,63	0,16	0,63	0,60
18	0,84	0,67	0,17	0,59	0,65
19	0,86	0,70	0,16	0,56	0,69
20	0,88	0,73	0,17	0,49	0,76
21	0,82	0,64	0,16	0,62	0,61
22	0,74	0,49	0,14	0,80	0,36
23	0,61	0,41	0,20	0,84	0,31
24	0,77	0,53	0,14	0,79	0,37
25	0,83	0,69	0,19	0,58	0,66
26	0,85	0,72	0,19	0,49	0,76
Questionnaire abrégé <sup>2</sup>					
1	0,83	0,79	s. o.	0,61	0,63
2	0,68	0,47	0,15	0,81	0,35
3	0,77	0,63	0,12	0,71	0,50
4	0,83	0,75	0,09	0,59	0,66
5	0,80	0,64	0,12	0,68	0,54
6	0,67	0,46	0,15	0,81	0,33
7	0,83	0,75	0,10	0,59	0,65
8	0,89	0,82	0,11	0,40	0,84

1. GFI = 0,88; RMSEA = 0,08;  $\chi^2 = 8599,7$ ; dl = 188 et p < 0,0001.

2. GFI = 0,97; RMSEA = 0,08;  $\chi^2 = 793,5$ ; dl = 19 et p < 0,0001.

L'analyse factorielle basée sur les huit énoncés du questionnaire abrégé montre qu'un seul facteur émerge. La valeur de la racine latente associée à ce facteur est de 4,98/8 et explique 62,3% de la variance totale. La fiabilité de la version abrégée est de 0,91. Ce taux est très satisfaisant. Ces résultats sont comparables à ce qui est observé pour la version originale. Le tableau 5 présente les pondérateurs associés au facteur principal. La structuration factorielle attendue en  $H_{1a}$  a été observée à partir des deux versions du questionnaire; seul un facteur principal a émergé.

### *L'analyse des commentaires*

Dans la version originale, 2081121 caractères ont été émis par 4525 répondants. Cela représente un ratio de 459 caractères par répondant, alors que 34% des répondants ( $n = 2345/6870$ ) n'expriment aucun commentaire. Au total, 16576 commentaires ont été émis, pour un ratio de 3,66 commentaires par répondant. Cependant, la longueur moyenne des commentaires est de 121,7 caractères dans la version originale. À chaque question, de 2,75% à 13,64% des répondants inscrivent un commentaire (voir Tableau 6). La moyenne est de 7,59% de répondants par question. Lorsqu'il y a des commentaires, la longueur moyenne varie entre 57,6 et 151,9 caractères, en fonction de l'énoncé. À titre d'exemple, pour l'énoncé 10, 937 répondants ont inscrit un commentaire, ce qui représente 13,64% des répondants; la longueur moyenne de leur commentaire est de 122 caractères, avec un écart-type de 106,4 caractères. La zone la plus utilisée pour s'exprimer est la zone générale à la fin du questionnaire. En tout, 43,92% (3017/6870) des répondants y ont inscrit un commentaire.

Tableau 6  
*Analyse des commentaires – Questionnaire original*

<b>Dimension</b>	<b>Item</b>	<b>n</b>	<b>%</b>	<b>Long.</b>	<b>É.-t.</b>
Contexte	1	468	6,81	94,1	71,4
	2	692	10,07	109,7	83,6
	3	244	3,55	90,9	86,2
	4	823	11,98	57,6	60,4
	5	401	5,84	75,3	68,0
Organisation/Clarté	6	374	5,44	90,0	75,7
	7	313	4,56	106,1	88,2
	8	594	8,65	116,2	116,8
	9	326	4,75	121,4	116,0
	10	937	13,64	122,0	106,4
	11	811	11,80	111,3	104,9
	12	653	9,51	84,0	74,7
	13	543	7,90	93,3	95,9
Dynamisme/Intérêt	14	635	9,24	76,4	69,2
	15	592	8,62	97,5	74,6
	16	417	6,07	110,8	107,7
	17	492	7,16	94,1	89,0
Interaction	18	505	7,35	95,7	80,2
	19	347	5,05	91,8	76,0
	20	213	3,10	94,9	88,1
Évaluation/Rétroaction	21	566	8,24	151,9	183,2
	22	760	11,00	107,6	91,7
	23	561	8,17	74,1	66,4
	24	637	9,27	108,3	107,2
Globale	25	466	6,78	96,0	73,1
	26	189	2,75	85,2	74,0
Générales	s. o.	3 017	43,92	223,6	225,8
<i>Total</i>		<i>16 576</i>	<i>8,9<sup>1</sup></i>	<i>121,7</i>	<i>108,9</i>

Item = le numéro de l'énoncé

n = le nombre de répondants qui ont émis un commentaire (N = 6 870)

% = le pourcentage (n/N) associé

Long. = la longueur moyenne des énoncés en nombre de caractères

É.-t. = l'écart-type

1. Ce nombre est obtenu comme suit : 16 576 répondants /  
 (6 870 répondants \* 27 possibilités de réponse) x 100

Dans la version abrégée, 1582194 caractères ont été écrits par 3792 répondants. Cela représente 417,2 caractères par répondant. Au total, 10876 commentaires ont été exprimés, ce qui représente un ratio de 2,87 commentaires par répondant. La longueur moyenne des commentaires est de 145,4 caractères dans la version abrégée (voir Tableau 7). Par contre, 47,7% (3465/7257) des répondants n'ont pas émis de commentaires, contre 34% dans la version originale.

Tableau 7  
*Analyse des commentaires – Questionnaire abrégé*

Question	Hiver 2015			
	n	%	Long.	É.-t.
Q1. Le cours est bien organisé.	1 228	16,9	153,1	149,7
Q2. Le plan de cours a été respecté.	585	8,1	124,7	114,9
Q3. L'enseignant maîtrise la matière.	1 158	16,0	119,0	114,2
Q4. L'enseignant démontre de bonnes habiletés d'enseignement.	1 656	22,8	155,9	170,7
Q5. L'enseignant évalue avec rigueur les apprentissages.	1 323	18,2	183,2	191,6
Q6. L'enseignant est disponible en dehors des heures de cours.	829	11,4	87,2	81,0
Q7. Le cours permet de réaliser des apprentissages significatifs.	998	13,8	148,9	140,3
Q8. Dans l'ensemble, je suis satisfait de ce cours.	972	13,4	164,0	176,7
Q9. Les conditions matérielles et temporelles du cours sont adéquates.	1 017	14,0	128,2	119,5
Q10. Le cours est pertinent dans le programme.	1 110	15,3	155,2	167,0
<i>Total</i>	<i>10 876</i>	<i>14,9</i>	<i>145,4</i>	<i>115,9</i>

n = le nombre de répondants

% = le pourcentage (n/N) associé

Long. = la longueur moyenne des énoncés en nombre de caractères

É.-t. = l'écart-type

Le pourcentage de répondants par question est cependant plus élevé dans la version abrégée; celui-ci se situe en moyenne à 14,9% des répondants, contre 8,9% dans la version originale.

Pour résumer, moins de répondants ont exprimé un commentaire dans la version abrégée; il y a également légèrement moins de caractères par répondant dans la version abrégée. Cependant, les messages sont un peu plus longs. Finalement, la proportion de répondants qui s'expriment à chaque question est supérieure à celle observée dans la version originale.

Les répondants ont donc exprimé leur opinion en tenant compte de la nouvelle réalité du questionnaire. Ils ont augmenté la longueur des commentaires et ils sont plus nombreux à s'exprimer à chaque question. Cela était attendu en vertu de l'hypothèse  $H_2$ .

Cependant, dans l'ensemble, les répondants sont moins nombreux à s'exprimer. L'absence d'une zone dédiée aux commentaires généraux à la fin du questionnaire explique probablement cette situation. Cette zone a été utilisée par plus de 43% des répondants dans la version originale. Il est donc prématuré de conclure que l'hypothèse  $H_3$  se confirme.

## ***Le comportement séquentiel***

### ***Analyse séquentielle***

Le tableau 8 présente la probabilité d'observer un type de réponse en fonction du type précédent pour les versions originale et abrégée. Plusieurs aspects de ces matrices peuvent être mis en exergue. Ainsi, la probabilité de répondre « tout à fait en accord » à un énoncé alors que la personne a répondu « tout à fait en accord » à l'énoncé précédent est de 84,6% pour la version originale. La probabilité de faire un autre choix n'est alors que de 15,4%. Il s'agit de la séquence la plus probable dans l'ensemble de la matrice.

Par ailleurs, la probabilité d'avoir un choix identique au précédent est toujours le choix le plus probable dans une colonne donnée de la matrice et très souvent le choix le plus probable dans une ligne de la matrice. Les exceptions sont pour les lignes 2 et 0.

Tableau 8  
*Comportement séquentiel*

*Probabilité d'observer un type de réponse en fonction  
du type précédent de réponse*

Questionnaire original					
Choix	0	1	2	3	4
0	0,287	0,054	0,105	0,259	0,295
1	0,020	0,425	0,233	0,210	0,112
2	0,014	0,097	0,279	0,408	0,202
3	0,010	0,029	0,116	0,577	0,269
4	0,005	0,006	0,024	0,119	0,846
Questionnaire abrégé					
Choix	0	1	2	3	4
0	0,194	0,050	0,111	0,316	0,329
1	0,047	0,413	0,212	0,211	0,116
2	0,041	0,114	0,266	0,361	0,218
3	0,028	0,029	0,105	0,504	0,333
4	0,012	0,005	0,024	0,137	0,822

Une autre caractéristique du comportement séquentiel est que plus un choix de réponse s'éloigne du choix précédent, plus la probabilité associée est faible. L'effet est donc décroissant. L'examen de la matrice de la version originale montre que les probabilités de répondre 3, 2, 1 ou 0 après avoir répondu 3 à un énoncé passent de 57,7% à 11,6% à 2,9% puis à 1% lorsque le choix s'éloigne du choix 3. Des patrons de réponses similaires existent pour les autres choix de réponse.

Il est également possible de constater que le comportement séquentiel s'applique aussi à l'option de non-réponse. Ainsi, globalement, la probabilité d'observer ce choix de réponse est de 1,33%. Cependant, cette probabilité augmente à 28,7% lorsque le répondant a choisi cette réponse lors de l'énoncé précédent.

De plus, la probabilité d'observer ce choix augmente lorsque la réponse précédente passe de 4 à 3, 2 et 1; ainsi, les probabilités d'observer ce choix passent de 0,5% à 1%, puis à 1,4% et, finalement, à 2% dans ces circonstances. La probabilité d'observer ce choix de réponse augmente donc avec l'insatisfaction croissante du répondant à la question précédente. En ce sens, ce choix de réponse témoigne d'une forme d'insatisfaction du répondant vis-à-vis du cours apprécié. Cette interprétation est singulière au présent article et n'est pas en vigueur à l'université, où ce choix de réponse reste ignoré et n'est pas considéré lors de la prise de décision.

Les mêmes caractéristiques du comportement séquentiel sont observées à partir de la version abrégée (voir Tableau 8). Ainsi, la probabilité de répondre « tout à fait en accord » alors que la personne a fait ce choix précédemment est celle qui est la plus élevée de la matrice, avec un taux de 82,2%. De plus, la probabilité de faire un choix identique au précédent est ici encore toujours le choix le plus dominant d'une colonne et très souvent le choix le plus probable d'une ligne, à l'exception des lignes 0 et 2. L'effet décroissant est aussi observé; plus un choix de réponse s'éloigne du choix précédent, plus la probabilité associée est faible. Cet effet s'observe également pour l'option de non-réponse. Comme nous l'avons mentionné pour la version originale, la probabilité d'observer ce choix augmente lorsque la réponse précédente passe de 4 à 3, 2 et 1; ainsi, les probabilités d'observer le choix « ne s'applique pas/je ne sais pas » passent de 1,2% à 2,8%, puis à 4,1% et, finalement, à 4,7% dans ces circonstances. La probabilité d'observer ce choix de réponse augmente donc avec l'insatisfaction croissante du répondant à la question précédente. Les pourcentages apparaissent plus élevés que dans la version originale, mais cette augmentation peut être attribuable au seul énoncé 6.

Il existe cependant une différence importante entre les matrices. En fait, toutes les probabilités de la diagonale sont inférieures dans la version abrégée par rapport à celles de la version originale, et ce, pour tous les choix de réponse (19,4% contre 28,7%; 41,1% contre 42,5%; 26,6% contre 27,9%; 50,4% contre 57,7%; et 82,2% contre 84,6%). Cela suggère que le comportement séquentiel est toujours bien présent dans la version abrégée, mais que celui-ci est légèrement moins prononcé que dans la version originale; le répondant a légèrement tendance à donner une réponse différente de la précédente, qui se situe donc en dehors de la diagonale. Cela constitue un léger avantage de la version abrégée.

### ***L'analyse factorielle confirmatoire***

Les paramètres de l'analyse factorielle confirmatoire sont présentés dans le tableau 5 pour les versions originale et abrégée. Le coefficient associé au facteur de mémorisation est de 0,165 et s'avère significatif ( $t = 61,16$ ;  $p < 0,001$ ) pour la version originale. Dans la version abrégée, ce coefficient est de 0,119 et s'avère tout aussi significatif ( $t = 21,83$ ;  $p < 0,001$ ). Le facteur de mémorisation apparaît cependant légèrement inférieur à celui de la version originale. Cela est cohérent avec la légère diminution du comportement séquentiel rapportée dans l'analyse précédente.

### ***L'option de non-réponse***

Dans l'ensemble de la version originale, l'option de non-réponse constitue environ 1% des choix de réponse. Ce choix a été enregistré à 1768 reprises, sur une possibilité de 178620 réponses ( $6870 \times 26$ ). Ce choix de réponse est sélectionné à tous les énoncés du questionnaire, avec une proportion qui varie entre 0,28% et 4,43% (voir Tableau 9).

Dans la version abrégée (voir Tableau 10), l'option de non-réponse est inférieure à 1% pour toutes les questions, sauf deux : les items 5 et 6. L'item 5 porte sur l'évaluation des apprentissages ; le taux est alors de 1,78%. Ce taux n'est pas problématique, car il est nettement inférieur à celui enregistré dans la version originale. L'item 6 mentionne que le professeur est disponible en dehors des heures de cours ; cet item apparaît ambigu, car le taux de l'option de non-réponse se situe à 11,62%. Cet item devrait faire l'objet d'une révision ; un item similaire dans la version originale, mais formulé différemment ne présente pas cette lacune.

Tableau 9  
*Fréquence de l'option de non-réponse – Questionnaire original*

Dimension	N° d'énoncé	Fréquence	%
Contexte	1	52	0,76
	2	113	1,64
	3	82	1,19
	4	62	0,90
	5	123	1,79
Organisation/Clarté	6	19	0,28
	7	30	0,44
	8	27	0,39
	9	34	0,49
	10	28	0,41
	11	20	0,29
	12	86	1,25
	13	44	0,64
Dynamisme/Intérêt	14	28	0,41
	15	44	0,64
	16	34	0,49
	17	75	1,09
Interaction	18	28	0,41
	19	86	1,25
	20	49	0,71
Évaluation/Rétroaction	21	42	0,61
	22	198	2,88
	23	84	1,22
	24	304	4,43
Globale	25	41	0,60
	26	35	0,51

Tableau 10  
*Fréquence de l'option de non-réponse en fonction des questions*  
*Questionnaire abrégé*

Question	Hiver 2015	
	Fréquence	%
Q1. Le cours est bien organisé.	35	0,48
Q2. Le plan de cours a été respecté.	49	0,68
Q3. L'enseignant maîtrise la matière.	46	0,63
Q4. L'enseignant démontre de bonnes habiletés d'enseignement.	37	0,51
Q5. L'enseignant évalue avec rigueur les apprentissages.	129	1,78
Q6. L'enseignant est disponible en dehors des heures de cours.	843	11,62
Q7. Le cours permet de réaliser des apprentissages significatifs.	49	0,68
Q8. Dans l'ensemble, je suis satisfait de ce cours.	42	0,58
Q9. Les conditions matérielles et temporelles du cours sont adéquates.	53	0,73
Q10. Le cours est pertinent dans le programme.	71	0,98

## Conclusion

L'objectif de cet article est de documenter le processus de validation d'une démarche abrégée d'évaluation de l'enseignement par les étudiants. Pour ce faire, les qualités psychométriques d'une version originale sont comparées à celles d'une version abrégée.

L'hypothèse principale ( $H_1$ ) considère que la version abrégée permet d'augmenter les taux de réponse des étudiants sans en changer les qualités métriques. Cette hypothèse n'est pas confirmée. Ainsi, bien que les taux de satisfaction des étudiants demeurent élevés et se situent à environ 86% dans les deux versions du questionnaire, les taux de réponse sont semblables dans les deux versions du questionnaire. Ils demeurent nettement inférieurs à ce qu'ils étaient lors de l'utilisation d'une version papier du même questionnaire. Il faut mentionner qu'il n'y a pas eu de promotion de la version abrégée et qu'en ce sens, les étudiants n'étaient pas informés du changement de questionnaire ni de ses avantages.

En contrepartie, plusieurs qualités psychométriques sont conservées. La dimensionnalité et la fiabilité du construit des deux versions sont comparables. La dimensionnalité des deux versions se présente sous la forme d'un facteur principal unique qui explique plus de 60% de la variance totale. L'hypothèse émise ( $H_{1a}$ ) est donc confirmée. La fiabilité est supérieure à 90% pour les deux versions. Ce résultat est conforme à ce qui est rapporté dans la documentation (Harvey & Hébert, 2012).

Deux hypothèses concernent l'expression des commentaires dans les deux versions du questionnaire. L'une considère qu'un étudiant qui a une opinion à exprimer le fera, et ce, peu importe le format du questionnaire ( $H_2$ ). Celle-ci suggère que le format du questionnaire n'influence pas le nombre ni la longueur des commentaires. L'autre hypothèse ( $H_3$ ) est que le format du questionnaire a une influence et que les questions jouent un rôle d'élicitation; ainsi, plus le nombre de questions est grand, plus le nombre de commentaires est important.

Les résultats suggèrent que les étudiants se sont adaptés au format du questionnaire et qu'ils ont exprimé leurs idées à propos de l'enseignement reçu. Cependant, l'adaptation n'est pas parfaite. Ainsi, la version abrégée génère légèrement moins de commentaires que la version originale. De

plus, un pourcentage moins élevé d'étudiants inscrit un commentaire; l'ensemble des commentaires (en nombre absolu de caractères) est donc légèrement moins long. La manière de répondre est différente. Les étudiants écrivent un peu moins de commentaires, mais la longueur des commentaires est légèrement plus longue.

Dans l'ensemble, cette diminution relative des commentaires peut s'expliquer par le fait qu'il n'y a pas de zone dédiée à l'inscription d'un commentaire général à la fin du questionnaire. Il apparaît donc propice d'ajouter une telle zone afin de susciter autant de commentaires dans chaque version du questionnaire. Par conséquent, il faut conclure qu'un étudiant qui a une opinion à exprimer le fera, mais le format du questionnaire a son importance. Les hypothèses  $H_2$  et  $H_3$  sont donc partiellement validées.

L'article permet également de documenter le comportement séquentiel des répondants. L'analyse séquentielle a montré que les répondants ont une forte propension à répéter la réponse antérieure. Cette propension varie entre 19% et 84% en fonction du type de réponse et du questionnaire. Cette propension est observée dans les deux versions du questionnaire, mais semble être moins prononcée dans la version abrégée; les répondants ont légèrement moins tendance à répéter le même choix d'une réponse à l'autre. La différence atteint même plus de 9% dans le cas de l'option de non-réponse. Il est positif de constater que la répétition de ce choix, qui est associée à de l'indécision du répondant ou à de l'ambiguïté des réponses, est moindre. Ces résultats sont cohérents avec l'hypothèse que l'ensemble du test génère moins d'incertitude de la part du répondant et qu'il existe moins d'ambiguïté quant aux questions que dans la version originale.

Cette diminution du comportement séquentiel est également observée à partir d'une analyse confirmatoire, où un modèle comprenant un facteur général et un facteur de mémorisation est utilisé (Green & Hershberger, 2000). Le facteur de mémorisation est légèrement inférieur dans la version abrégée.

Le présent article permet aussi de mieux comprendre l'utilisation de l'option de non-réponse dans le contexte de l'EEE. Premièrement, ce type de réponse représente environ 1% des réponses dans la version originale. Il est réparti dans l'ensemble des questions, mais est particulièrement fréquent dans les questions portant sur l'évaluation des travaux ou sur la

rétroaction qui les concerne. Ce type de réponse peut alors représenter 5% des réponses. Dans la version abrégée, ce choix de réponse est inférieur à 1% pour toutes les questions, sauf une. Les changements qui ont été faits au questionnaire ont donc permis de réduire l'ambiguïté des items ou l'ambivalence des répondants à une exception près, soit l'item 6. Cet item mentionne que le professeur est disponible en dehors des heures de cours ; il apparaît ambigu, à la lumière du taux de l'option de non-réponse. Il est possible que les répondants n'aient pas eu à vérifier la disponibilité de leur professeur en dehors des heures de cours et considèrent que l'item ne s'applique pas ou qu'ils ne savent pas si leur professeur était disponible ou non en dehors des heures de cours.

Ces résultats liés à l'expression des commentaires, au comportement séquentiel qui est moindre et à l'utilisation moins prononcée de l'option de non-réponse tendent à suggérer que la version abrégée et sa structure réduite n'altèrent pas la fonction d'élicitation associée à la structuration séquentielle du questionnaire original.

Compte tenu de ces caractéristiques, la version abrégée du questionnaire peut être utilisée en substitution à la version originale en conservant les qualités psychométriques, ce qui rejoint d'autres auteurs (Dolmans & Ginns, 2005).

Il est également important de préciser qu'aux fins de la présente recherche, l'analyse du contenu des commentaires n'a pas été réalisée. L'argument est que, étant donné que les qualités psychométriques des instruments sont comparables, il n'y a aucune raison de penser que la nature des commentaires a changé dans la version abrégée. De plus, l'analyse du contenu des commentaires de plus de 8 000 répondants nécessite une démarche exhaustive, qui apparaît clairement irréaliste. La sélection d'un sous-ensemble de répondants même au hasard ouvre la possibilité d'un biais d'échantillonnage et d'une observation de différences artificielles. Pour ces raisons, l'analyse du contenu des commentaires n'a pas été réalisée.

Réception : 6 juillet 2015

Version finale : 7 juin 2016

Acceptation : 9 juin 2016

## Annexe

---

### Questionnaire d'appréciation de l'enseignement en classe ordinaire<sup>1</sup>

Afin d'améliorer la qualité des enseignements, vous avez l'occasion d'apprécier la façon dont l'enseignement s'est donné, le contexte dans lequel le cours a eu lieu et sa pertinence dans votre programme. Nous vous invitons à accompagner vos réponses de commentaires liés à des points forts ou à des points faibles que vous avez identifiés et, le cas échéant, de suggestions visant certaines améliorations. Merci de votre participation !

*L'appréciation de l'enseignement débute maintenant :*

- 
- N° 1** **Le cours est bien organisé** (préparation, contenu et structure du cours, charge de travail).
- 
- N° 2** **Le plan de cours convenu a été respecté** (modalités d'évaluation, calendrier des rencontres, échéancier pour la remise des travaux).
- 
- N° 3** **L'enseignant(e) maîtrise la matière** (niveau de connaissances, capacité à approfondir les thèmes du cours).
- 
- N° 4** **L'enseignant(e) démontre de bonnes habiletés d'enseignement** (dynamisme, clarté des explications, capacité à lier théorie et pratique, à synthétiser la matière, à s'adapter aux imprévus, à interagir avec les étudiants, formules pédagogiques utilisées).
- 
- N° 5** **L'enseignant(e) évalue avec rigueur les apprentissages** (moyens d'évaluation, clarté des consignes, critères d'évaluation, pondération des travaux et des examens, délai et qualité de la correction, richesse de la rétroaction).
- 
- N° 6** **L'enseignant(e) est disponible en dehors des heures de cours** (périodes de consultation suffisantes, facilement joignable, délai raisonnable de réponse).
- 
- N° 7** **Le cours permet de réaliser des apprentissages significatifs** (connaissances, compétences, aptitudes acquises, progression personnelle, niveau du cours ou autres retombées du cours).
- 
- N° 8** **Dans l'ensemble, je suis satisfait(e) de ce cours.**

#### Section non compilée dans l'appréciation de l'enseignant(e)

*Ces deux questions portent sur le contexte et la pertinence du cours. Merci d'y répondre.*

- 
- N° 9** **Les conditions matérielles et temporelles du cours sont adéquates** (horaire du cours, disponibilité des notes de cours et des livres obligatoires, groupe d'étudiants, salle de classe et équipements).
- 
- N° 10** **Le cours est pertinent dans le programme** (objectifs du cours, contenu, place dans le programme).
- 

1. Afin d'économiser de l'espace, l'échelle de Likert à quatre modalités et la zone destinée aux commentaires écrits ne sont pas reproduites ici; seul l'énoncé de chaque question apparaît.

## NOTE

1. Le questionnaire utilisé en classe ordinaire se distingue de ceux utilisés en classe virtuelle, dans les cours laboratoires, pour l'enseignement individualisé ou pour les cours stages, mais la même logique s'applique pour ces autres situations d'enseignement.

## RÉFÉRENCES

- American Educational Research Association (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.
- André, N., Loye, N. & Laurencelle, L. (2015). La validité psychométrique: un regard global sur un concept centenaire, sa genèse, ses avatars. *Mesure et évaluation en éducation*, 37(3), 125-148. Repéré à: [https://www.researchgate.net/publication/281639436\\_La\\_validite\\_psychometrique\\_un\\_regard\\_global\\_sur\\_le\\_concept\\_centenaire\\_sa\\_genese\\_ses\\_avatars](https://www.researchgate.net/publication/281639436_La_validite_psychometrique_un_regard_global_sur_le_concept_centenaire_sa_genese_ses_avatars)
- Bernard, H. (2011). *Comment évaluer, améliorer, valoriser l'enseignement supérieur*. Bruxelles: De Boeck.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111, 1061-1071. doi: 10.1037/0033-295X.111.4.1061
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.
- Carley-Baxter, L. R., Peytchev, A. A., & Black, M. L. (2010, May). *Effect of questionnaire structure on nonresponse and measurement error: Sequential vs. grouped placement of filter questions*. Paper presented at the annual conference of the American Association for Public Opinion Research, Chicago (IL).
- Cizek, G. J. (2012). Defining and distinguishing validity: Interpretations of score meaning and justifications of test use. *Psychological Methods*, 17, 31-43. doi: 10.1037/a0026975
- d'Apollonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist*, 52(11), 1198-1208. doi: 10.1037/0003-066X.52.11.1198
- Detroz, P. (2008). L'évaluation des enseignements par les étudiants: état de la recherche et perspectives. *Revue française de pédagogie*, 165, 117-135. doi: 10.4000/rfp.1165
- Dolmans, D. H. J. M., & Ginns, P. (2005). A short questionnaire to evaluate the effectiveness of tutors in PBL: Validity and reliability. *Medical Teacher*, 27(6), 534-538. doi: 10.1080/01421590500136477
- Donovan, J., Mader, C. E., & Shinsky, J. (2006). Constructive student feedback: Online vs. traditional course evaluations. *Journal of Interactive Online Learning*, 5(3), 283-296.
- Feick, L. F. (1989). Latent class analysis of survey questions that include don't know responses. *Public Opinion Quarterly*, 53, 525-547. doi: 10.1086/269170

- Fontaine, S. (2009). Des expériences actuelles d'évaluation des enseignements vers des démarches adaptées aux 2<sup>e</sup> et 3<sup>e</sup> cycles. Dans M. Romainville & C. Coggi (dir.), *L'évaluation de l'enseignement par les étudiants : approches critiques et pratiques innovantes* (pp. 123-144). Bruxelles: De Boeck.
- Gal, Y., & Gal, A. (2015). Knowledge bias by utilizing the wording on feedback questionnaires: A case study of an Israeli college. *Journal of the Knowledge Economy*, 1-18.
- Green, S. B., & Hershberger, S. L. (2000). Correlated errors in true score models and their effect on coefficient Alpha. *Structural Equation Modeling: A Multidisciplinary Journal*, 7(2), 251-270. doi: 10.1207/S15328007SEM0702\_6
- Harvey, L. & Hébert, M.-H. (2012). Évaluation de la qualité de l'enseignement par les étudiantes et étudiants : qualités psychométriques et comparaison des conditions de passation. *Mesure et évaluation en éducation*, 35(3), 31-60. doi: 10.7202/1024669ar
- Holman, R., Glas, C. A. W., Lindeboom, R., Zwinderman, A. H., & de Hann, R. J. (2004). Practical methods for dealing with "not applicable" item responses in the AMC Linear Disability Score project. *Health and Quality of Life Outcomes*, 2(29). doi: 10.1186/1477-7525-2-29
- Huxham, M., Laybourn, P., Cairncross, S., Gray, M., Brown, N., Goldfinch, J., & Earl, S. (2008). Collecting student feedback: A comparison of questionnaire and other methods. *Assessment & Evaluation in Higher Education*, 33(6), 675-686. doi: 10.1080/02602930701773000
- Hwang, J., Petrolia, D. R., & Interis, M. G. (2014). Consequentiality and opt-out responses in stated preference surveys. *Agricultural and Resource Economics Review*, 43(3), 471-488. Retrieved from: [https://www.researchgate.net/publication/269036653\\_Consequentiality\\_and\\_Opt-out\\_Responses\\_in\\_Stated\\_Preference\\_Surveys](https://www.researchgate.net/publication/269036653_Consequentiality_and_Opt-out_Responses_in_Stated_Preference_Surveys)
- Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Hanemann, W. M., Kopp, R. J., ... Conaway, M. (2002). The impact of "no opinion" response options on data quality: Non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly*, 66, 371-403. Retrieved from: <https://pprg.stanford.edu/wp-content/uploads/2002-The-impact-of-no-opinion-response-options-on-data-quality-N.pdf>
- Marsh, H. W., & Bailey, H. (1993). Multidimensional students' evaluations of teaching effectiveness: A profile analysis. *Journal of Higher Education*, 64(1), 1-18. doi: 10.2307/2959975
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749. doi: 10.1037//0003-066X.50.9.741
- Raykov, T. (2004). Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. *Behavior Therapy*, 35, 299-331. doi: 10.1016/S0005-7894(04)80041-8
- Redding, C. A., Maddock, J. E., & Rossi, J. S. (2006). The sequential approach to measurement of health behavior construct: Issues in selecting and developing measures. *Californian Journal of Health Promotion*, 4(1), 83-101. Retrieved from: [http://www.cjhp.org/Volume4\\_2006/Issue1/83-101-redding.pdf](http://www.cjhp.org/Volume4_2006/Issue1/83-101-redding.pdf)

- Saupe, J. L., & Eimers, M. T. (2012, June). *Alternative estimates of the reliability of college grade point averages*. Paper presented at the 52nd annual forum of the Association for Institutional Research, New Orleans, LA.
- Thorpe, S. W. (2002). *On-line student evaluation of instruction: An investigation of non-response bias*. Paper presented at the 42nd annual forum of the Association for Institutional Research, Toronto, Canada.