

Évaluation de la qualité de l'enseignement par les étudiantes et étudiants

Qualités psychométriques et comparaison des conditions de passation

Léon Harvey and Marie-Hélène Hébert

Volume 35, Number 3, 2012

URI: <https://id.erudit.org/iderudit/1024669ar>

DOI: <https://doi.org/10.7202/1024669ar>

[See table of contents](#)

Publisher(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (print)

2368-2000 (digital)

[Explore this journal](#)

Cite this article

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. <https://doi.org/10.7202/1024669ar>

Article abstract

Students' evaluations of teaching effectiveness are used for academic and formative purposes, but are also sometimes evoked, in a summative way, for some faculty career decisions. In the present study, two main objectives are pursued. The first objective is to determine the psychometric properties of the instrument in use at the Université du Québec à Rimouski. With the advent of information and communication technologies, the conditions in which the evaluations take place have been changed from paper to online processes. so, the second objective is to verify the impacts of these online vs paper processes on the psychometric properties of the instrument. The data originated from the evaluations gathered during the 2007-2008 (20,245 evaluations) and 2010-2011 (16,432 evaluations) academic years. The results attest that a hierarchical model better describes the data than alternative models, that online vs paper processes have an influence on response rates and, in some cases, on the satisfaction level with teaching, but they have no impact on the reliability and the dimensionality of the evaluations. The implications of these results are discussed.

Évaluation de la qualité de l'enseignement par les étudiantes et étudiants : qualités psychométriques et comparaison des conditions de passation

Léon Harvey

Marie-Hélène Hébert

Université du Québec à Rimouski

MOTS CLÉS : évaluation de l'enseignement, modalités, en ligne vs papier, validité

Les évaluations de l'enseignement par les étudiants et étudiantes sont utilisées à des fins académiques et formatives, mais sont également évoquées d'une manière sommative lors de décisions liées à la carrière professorale. Dans le cadre de la présente recherche, deux objectifs sont poursuivis. Le premier objectif est de déterminer les qualités psychométriques de l'instrument utilisé à l'Université du Québec à Rimouski. Avec l'avènement des technologies de l'information et de la communication, les conditions dans lesquelles se déroulent ces évaluations ont été modifiées. Le second objectif est de vérifier les impacts de la modalité de passation en ligne vs papier sur les qualités psychométriques de l'instrument. Le corpus de données est constitué des évaluations réalisées durant les années 2007-2008 (20 245 évaluations) et 2010-2011 (16 432 évaluations). Les résultats attestent qu'un modèle hiérarchique décrit mieux la structuration du questionnaire que des modèles alternatifs, que la modalité de passation a une influence sur le taux de réponses et, dans certains cas, sur le taux de satisfaction des enseignants, mais qu'elle n'influence ni la fiabilité ni la dimensionnalité des évaluations. Les implications de ces résultats sont discutées.

KEY WORDS: students' evaluations of teaching effectiveness, online vs paper processes, validity

Students' evaluations of teaching effectiveness are used for academic and formative purposes, but are also sometimes evoked, in a summative way, for some faculty career decisions. In the present study, two main objectives are pursued. The first objective is to determine the psychometric properties of the instrument in use at the Université du Québec à Rimouski. With the advent of information and communication technologies, the conditions in which the evaluations take place

have been changed from paper to online processes. So, the second objective is to verify the impacts of these online vs paper processes on the psychometric properties of the instrument. The data originated from the evaluations gathered during the 2007-2008 (20,245 evaluations) and 2010-2011 (16,432 evaluations) academic years. The results attest that a hierarchical model better describes the data than alternative models, that online vs paper processes have an influence on response rates and, in some cases, on the satisfaction level with teaching, but they have no impact on the reliability and the dimensionality of the evaluations. The implications of these results are discussed.

PALAVRAS-CHAVE: *avaliação do ensino, modalidades, online vs. papel, validade*

As avaliações do ensino pelos estudantes são utilizadas para fins acadêmicos e formativos, mas igualmente evocados de uma maneira sumativa no momento das decisões ligadas à carreira docente. No quadro da presente investigação, são seguidos dois objetivos. O primeiro objetivo é determinar as qualidades psicométricas do instrumento utilizado na Universidade do Quebeque em Rimouski. Com o advento das tecnologias da informação e da comunicação, as condições nas quais se desenvolvem estas avaliações foram modificadas. O segundo objetivo é verificar os impactos da modalidade de administração online vs. papel sobre as qualidades psicométricas do instrumento. O corpo de dados é constituído por avaliações realizadas durante os anos 2007-2008 (20,245 avaliações) e 2010-2011 (16,432 avaliações). Os resultados atestam que um modelo hierárquico descreve melhor a estruturação dos questionários que os modelos alternativos, que o modo de administração tem uma influência sobre a taxa de respostas e, em certos casos, sobre a taxa de satisfação do ensino, mas que não tem influência nem na fiabilidade nem na dimensionalidade das avaliações. As implicações destes resultados são discutidas.

Note des auteurs – Toute correspondance peut être adressée comme suit : Léon Harvey, Unité des sciences de l'éducation de Rimouski, Université du Québec à Rimouski, 300 allée des Ursulines, Rimouski (Québec), G5L 3A1, téléphone : 418 723-1986, poste 1758, ou par courriel à l'adresse suivante : [Leon_Harvey@uqar.ca].

Introduction

L'évaluation de l'enseignement par les étudiants (EEE) désigne « les procédures par lesquelles les étudiants prennent part au jugement porté sur des cours ou des enseignements spécifiques, attribués à des enseignants ou à des équipes d'enseignants au sein d'un programme ou d'une formation » (Romainville & Coggi, 2009, p. 13). Actuellement, la grande majorité des universités aux États-Unis (Dommeyer, Baum, Hanna, & Chapman, 2004), au Canada et en Europe évaluent régulièrement la qualité de l'enseignement à partir de l'opinion des étudiantes et des étudiants inscrits aux activités d'enseignement. Les résultats de ces évaluations sont utilisés à des fins académiques mais sont également évoqués lors de décisions liées à la carrière professorale (Donovan, Mader, & Shinsky, 2006 ; Fontaine ; 2009 ; Younès, 2009). Ils sont, par conséquent, une source de controverses dans les milieux universitaires (Dommeyer et al., 2004 ; Romainville & Coggi, 2009 ; UQAR, 2011a). Deux aspects parmi les plus controversés concernent le construit qui est effectivement mesuré par ces évaluations et la baisse du taux de réponses lorsque l'évaluation est réalisée *en ligne* plutôt qu'en classe en utilisant un format *papier*.

Cette étude vise deux objectifs. Le premier consiste à déterminer les propriétés psychométriques d'un instrument de mesure de la qualité de l'enseignement par les étudiantes et les étudiants. Le second vise à déterminer l'impact des conditions de passation sur ces propriétés lorsque l'instrument est administré selon deux modalités : *papier* et *en ligne*.

Évaluation de la qualité de l'enseignement

Dimensionnalité

La documentation sur l'évaluation de la qualité de l'enseignement par les étudiants et les étudiantes est vaste et remonte jusqu'aux années 1920 (Marsh & Bailey, 1993). Sur le plan théorique, il n'y a pas d'unanimité sur la façon de définir ce qu'est un enseignement de qualité et sur la façon de le mesurer. Romainville (2009) soutient qu'il n'existe pas de consensus, ni de modèle sur ce qu'est un bon enseignement. Abrami, d'Apollonia, et Rosenfield (2007) mentionnent qu'il existe des définitions axées sur le produit de l'enseignement, sur les processus liés à l'enseignement et sur une mixture des deux pre-

mières définitions. Du point de vue du produit, un enseignement de qualité est celui qui permet de produire chez les étudiantes et les étudiants des changements significatifs sur le plan des apprentissages réalisés, que ceux-ci relèvent des domaines cognitif, affectif ou psychomoteur. Du point de vue des processus, un enseignement de qualité met l'emphase sur des actes liés à l'enseignement plutôt que sur les conséquences de ces actes. Ces actes sont alors liés à la préparation du cours, à la clarté des objectifs, aux interactions avec le groupe, au dynamisme, à l'évaluation et ainsi de suite. Finalement, l'approche hybride processus-produit explicite la relation entre les actes d'enseigner et les produits attendus.

Sur le plan empirique, Marsh (1983, 1984 ; Marsh & Bailey, 1993) affirme que les évaluations de la qualité de l'enseignement par les étudiantes et les étudiants sont multidimensionnelles, fiables et stables, fonction de la personne qui enseigne et non une propriété du cours qui est enseigné, valides face à une variété d'indicateurs d'enseignement de qualité, non affectées par des biais hypothétiques et utiles à différents acteurs: aux professeurs¹ à titre de rétroaction à propos de leur enseignement, aux étudiants en tant qu'outil de sélection de cours, aux administrateurs à des fins de décisions ainsi qu'aux chercheurs.

Le présent article répertorie quelques instruments de mesure de la qualité de l'enseignement par les étudiantes et les étudiants (voir tableau 1). L'un des outils standardisés parmi les plus cités et complexes est le *Students' Evaluation of Educational Quality* (SEEQ). Cet instrument est composé de neuf dimensions et de 35 items (Marsh, 1983, 1984). À l'opposé, Monsen, Woo, Mahan, et Miller (2005) rapportent qu'une université (Yale Law) a optimisé son processus d'évaluation en élaborant une version abrégée de son instrument. Celle-ci, composée de huit items, a permis de rehausser le taux de participation des étudiantes et des étudiants, avec l'utilisation conjointe d'une mesure coercitive. Entre ces extrêmes, plusieurs instruments (Byrne & Flood, 2003 ; Cohen, 2005 ; Hess et al., 2005 ; Layne, DeCristoforo, & McGinty, 1999 ; Thorpe, 2002 ; Toland & De Ayala, 2005 ; Wong & Fitzsimmons, 2008) mesurent entre deux et cinq dimensions et sont constitués de 12 à 25 items. Il existe donc une certaine variabilité sur le construit qui est mesuré. La plupart de ces instruments mesurent des dimensions primaires (de premier ordre) qui se rapprochent, dans certains cas, des trois rôles de l'enseignant retenus par Feldman (1976) et repris dans les travaux d'Abrami (Abrami et al., 2007 ; Abrami, Rosenfield, & Dedic, 2007 ; d'Apollonia & Abrami, 1997). Ces rôles

sont l'enseignement de la matière du cours (*delivery of instruction*), la facilité d'interagir avec les étudiantes et les étudiants (*facilitation of interactions*), et l'évaluation des apprentissages (*evaluation of student learning*).

Tableau 1
Qualités psychométriques de quelques instruments utilisés pour l'EEE

<i>Études</i>	<i>ND^a</i>	<i>NI^b</i>	<i>Corrélations</i> (Dimensions)	<i>Alpha</i> (Dimensions)	<i>Dimensions</i>
Feldman (1976)	19	nd ^c	nd	nd	19 dimensions qui se regroupent en 3 rôles de l'enseignant
Marsh (1983)	9	35	nd	0,87 à 0,98	Apprentissage, Enthousiasme, Organisation, Interaction, Rapport individuel, Matière couverte, Examens, Travaux, Charge de travail
Hess et al. (2005)	3	15	0,52 à 0,63	0,84 à 0,87	Évaluation, Communication, Conception
Thorpe (2002)	4	23	nd	nd	Étudiant, Cours, Professeur, Global
Cohen (2005)	2	12	Nulles	nd	Cours, Professeur
Toland & De Ayala (2005)	3	25	nd	nd	Cours, Interaction, Évaluation
Wong & Fitzsimmons (2008)	3	19	0,80 à 0,87	0,91 à 0,93	Attributs personnels, Facilitation de l'apprentissage, Rétroactions
Layne et al. (1999)	3	14	Nulles	0,95 à 0,99	Préparation et présentation du cours, Interactions, Évaluation
Byrne & Flood (2003)	5	23	nd	0,66 à 0,78	Enseignement, Buts, Charge de travail, Évaluation, Global
UQAR	5(6)	21(26)	—	—	(Contexte), Organisation, Dynamisme, Interaction, Évaluation, Appréciation générale
Yale Law	nd	8	nd	nd	nd

^a ND = Nombre de dimensions

^b NI = Nombre d'items

^c nd = Non disponible

À l'Université du Québec à Rimouski (UQAR), le questionnaire mesure six dimensions. Il s'insère dans une perspective hybride, basée à la fois sur l'évaluation des produits et des processus. Il est composé de 26 items, dont 19 sont relatifs au processus lié à l'acte d'enseigner et deux sont liés au produit de l'enseignement. Parmi les 26 items se trouvent aussi des indicateurs (cinq items) de la qualité générale de l'environnement dans lequel se déroulent les apprentissages (horaire, matériel pédagogique, salle de cours, etc.). Il faut pré-

ciser qu'à l'UQAR, ces éléments de contexte sont indépendants de la ressource professorale. Des items de ce type sont aussi présents dans d'autres instruments (Fontaine, 2009 ; Thorpe, 2002). À l'UQAR, ils balisent la décision prise par l'institution au sujet de la ressource professorale, mais ne sont pas considérés dans le sommatif du test.

L'absence de consensus quant aux dimensions du construit qui est mesuré pose cependant un problème (d'Apollonia & Abrami, 1997) lorsque les résultats des évaluations sont utilisés à des fins décisionnelles. Marsh et Dunkin (1992) précisent d'ailleurs que la plupart des instruments sont basés sur une mixture de considérations logiques et pragmatiques plutôt que sur un construit théorique déterminé. Selon Marlin (1990, cité par Poissant, 1996), il existerait plusieurs milliers de questionnaires aux dimensions multiples et variées, dont le processus de construction et la qualité de l'évaluation qui en découle varient.

Abrami et ses collaborateurs (2007) remettent en question la pertinence d'une évaluation sommative basée sur des ensembles hétérogènes de dimensions spécifiques plus ou moins corrélées. D'Apollonia et Abrami (1997) rapportent que les matrices de corrélation de plusieurs instruments considérés multidimensionnels, dont le SEEQ, ont été réanalysées au moyen d'une analyse factorielle exploratoire. Pour l'ensemble des instruments, les résultats ont montré l'existence d'une composante principale forte qui explique entre 30 % et 60 % de la variance des dimensions. En particulier, 31 des 35 items du SEEQ saturent fortement ($\lambda > 0,60$) dans cette composante principale. Ce n'est qu'à la suite d'une rotation oblique que l'importance des neuf dimensions spécifiques corrélées apparaît. La thèse de la multidimensionnalité est alors favorisée par le modèle mathématique privilégié (une rotation oblique des composantes de l'analyse factorielle) et le choix d'inclure des items spécifiques lors de la conception du questionnaire.

Afin de consolider leur thèse d'une composante principale forte, ces auteurs (voir d'Apollonia et Abrami, 1997) ont utilisé des techniques de méta-analyses à partir de 458 items différents en provenance de 17 instruments. Outre les composantes primaires, un facteur de second ordre qui explique 63 % de la variance de la qualité de l'enseignement a été relevé. Il constitue ce qu'ils appellent une *Compétence générale en enseignement (General Instructional Skill)*. Mentionnons qu'une analyse de second ordre se distingue par le fait qu'elle explique les intercorrélations entre les dimensions primaires à partir de la présence d'un facteur hiérarchique.

Ces études sur la dimensionnalité de l'EEE sont pertinentes et s'insèrent dans l'ensemble du processus qui consiste à vérifier la validité des instruments utilisés. La *validité* est un concept unitaire. C'est un jugement qui est porté sur l'ensemble des preuves accumulées à propos d'un instrument et qui détermine si les interprétations réalisées à partir de celui-ci sont conformes à l'intention visée (American Educational Research Association, 1999 ; Laveault & Grégoire, 2002). Le concept de validité de construit (Zaltman, Pinson, & Angelman, 1973) fait référence au degré auquel une opérationnalisation (un instrument) permet de mesurer le concept qu'elle est supposée représenter. La documentation (voir Laveault & Grégoire, 2002) utilise aussi l'expression *validité de concept* comme synonyme et considère que les procédures associées donnent un sens aux scores obtenus au test en lien avec le modèle théorique que l'on souhaite évaluer. Linn et Miller (2005) spécifient qu'il s'agit de déterminer le degré auquel la performance à un test peut être interprétée en termes d'un ou de plusieurs construits. Ce degré est établi à partir de l'adéquation entre les structures théorique et empirique (De Ketele & Gérard, 2005 ; Louis, Jutras, & Hensler, 1996) de l'objet mesuré. L'utilisation de l'analyse factorielle confirmatoire permet de réaliser cette opération.

Fiabilité

Quant à la fiabilité, la documentation suggère qu'en général les instruments utilisés pour mesurer la qualité de l'enseignement sont fiables (*e.g.*, Marsh & Bailey, 1993). La fiabilité fait référence à la précision et à la consistance interne d'un instrument, c'est-à-dire à la proportion de variance attribuable au score vrai par rapport au total de la variance (Brown, 2006 ; Brunner, Nagy, & Wilhelm, à paraître ; Raykov, 2004). Elle permet de déterminer jusqu'à quel point un score est affecté par la présence de sources d'erreur. Dans la théorie classique des tests, l'alpha de Cronbach est utilisé sur les items qui composent chacune des dimensions et sur l'ensemble des items d'un instrument. Cet indice est celui qui est le plus régulièrement rapporté (voir tableau 1).

Dans la perspective où un modèle factoriel confirmatoire extrait des facteurs latents dont certains peuvent être de second ordre ou hiérarchiques, la documentation (Brown, 2006 ; Raykov, 2004) recommande l'utilisation d'indices appelés *oméga* (ω) et *oméga hiérarchique* (ω_h) (Brunner et al., à paraître). Ces indices témoignent de la proportion de variance expliquée par l'ensemble des items (ω), chacun des facteurs spécifiques (ω) ainsi que le facteur hiérarchique (ω_h). La présence d'un facteur hiérarchique peut changer considérablement la fiabilité des dimensions spécifiques d'un instrument par rap-

port à celle obtenue à partir d'un modèle de premier ordre. Dans cette perspective, la fiabilité des dimensions rapportées dans la documentation (voir tableau 1) peut être une surestimation.

Conditions de passation

Par ailleurs, les conditions dans lesquelles se déroule l'évaluation de la qualité de l'enseignement peuvent influencer la validité observée. Or, depuis la fin des années 90, avec l'avènement des technologies de l'information, les conditions dans lesquelles se déroulent les évaluations ont considérablement évolué et plusieurs universités mettent en place des dispositifs d'évaluation *en ligne*. Classiquement, les évaluations sont réalisées dans le contexte de la classe à partir d'instruments *papier* dans des conditions où il peut exister dans la procédure d'administration (Fontaine, 2009) des biais imputables à la captivité relative des étudiants et des étudiantes. L'avantage principal de l'évaluation *en ligne* comparativement à la modalité *papier* (Donovan et al., 2006; Reid, 2001) est qu'elle représente une économie importante pour ce qui est du temps et des ressources nécessaires à la gestion de l'opération. De plus, l'évaluation peut être réalisée lorsque l'étudiante ou l'étudiant y est disposé. Des études américaines et françaises rapportent que plus de 88,4% des étudiants préfèrent la modalité *en ligne* plutôt que *papier* (Donovan, Mader, & Shinsky, 2007) ou se disent *satisfaits* à 83% du dispositif *en ligne* (Younès, 2009). Cependant, le passage du mode *papier* au mode *en ligne* demeure parfois perçu comme une menace à la validité et à la fiabilité observées (Grim Fidelman, 2007). Crews et Curtis (2011) rapportent d'ailleurs que 43% des professeurs préfèrent l'évaluation *papier* plutôt qu'*en ligne* et certaines universités rejettent cette dernière modalité (UQAR, 2011a).

À ce sujet, les études comparatives entre les modalités de passation *papier* et *en ligne* suggèrent que le changement de modalité n'affecte pas le taux de satisfaction moyen de la qualité de l'enseignement (Cates, 1993; Dommeyer et al., 2004; Layne et al., 1999). Sur le plan de la dimensionnalité, Layne et al. (1999) rapportent que la même structure factorielle émerge sous les deux modes de passation. Cates (1993) note un niveau de fiabilité similaire entre les modalités *papier* et *en ligne* lorsqu'estimé au moyen de l'alpha de Cronbach. Par contre, un taux de participation plus faible est rapporté (Dommeyer et al., 2004; Layne et al., 1999) et est considéré comme étant la principale limite à l'utilisation de procédures informatisées (Liegle & McDonald, 2004). Afin

d'apporter un correctif à ce faible taux, certaines universités mettent alors en place des mesures incitatives ou coercitives afin d'influencer à la hausse le taux de participation (Crews & Curtis, 2011).

Objectifs et hypothèses de la recherche

Considérant ce qui a été exposé jusqu'à présent et soucieuse de maintenir un processus d'évaluation valide, l'Université du Québec à Rimouski, qui utilise depuis 2009 une version *en ligne* de son questionnaire, a créé en 2011 (UQAR, 2011b) un comité 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.

Dans le contexte du présent article, deux objectifs principaux sont poursuivis. Le premier consiste à déterminer les qualités psychométriques du questionnaire utilisé à l'université pour les cours offerts en mode régulier. En particulier, il s'agit de vérifier comment se structure le questionnaire. Se structure-t-il effectivement tel qu'attendu? Une structure hiérarchique ou un facteur de second ordre peuvent-ils être relevés? Quelle est alors la fiabilité des dimensions?

Le second objectif est de vérifier l'impact de la modalité de passation (*en ligne* vs *papier*) sur les qualités psychométriques de l'instrument.

Quatre hypothèses peuvent être énoncées en lien avec les deux objectifs :

Hypothèse 1 – Taux de réponses. Les modalités de passation auront un effet sur les taux de réponses. Plus spécifiquement, les taux de réponses de la version *en ligne* seront inférieurs à ceux de la version *papier*.

Hypothèse 2 – Taux de satisfaction. Les modalités de passation n'auront pas d'effet sur les taux de satisfaction de la qualité de l'enseignement. Les taux de satisfaction aux items, aux dimensions et au total pour les versions *en ligne* et *papier* seront identiques.

Hypothèse 3 – Fiabilité (alpha de Cronbach). Les modalités de passation n'auront pas d'effet sur la fiabilité de l'instrument. Les indicateurs de fiabilité aux dimensions et au total pour les versions *en ligne* et *papier* seront identiques.

Hypothèse 4 – Dimensionnalité. Un modèle factoriel de second ordre ou hiérarchique permettra de décrire la structure du questionnaire mieux qu'un modèle de premier ordre, et ce, sous les deux modes de passation. La structure factorielle sera similaire pour les versions *en ligne* et *papier*.

Il est à noter que les hypothèses 2 et 3 prévoient une absence de différence entre les modalités de passation. Cette formulation respecte l'absence de différences généralement observée dans la documentation.

Méthode

Questionnaire

Le questionnaire utilisé à l'UQAR pour l'EEE a été adopté en 1978, soit quelques années après la création de l'université. À l'époque, le questionnaire est générique dans le sens où il est appliqué à toutes les formes d'enseignement. En 1997, il est révisé et cinq versions adaptées aux contextes pédagogiques des cours sont produites. Ainsi, des versions appropriées pour évaluer les cours réguliers, les cours à distance, les stages, les laboratoires et le tutorat sont maintenant en usage. À notre connaissance, aucune étude de validité ni de fiabilité n'a été réalisée sur l'une ou l'autre des versions. Seule celle utilisée pour évaluer les cours réguliers est inspectée dans le cadre du présent article.

À travers ce questionnaire, six dimensions sont évaluées par les étudiants au départ de 26 items. Elles se déclinent comme suit : *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 type Likert à cinq modalités, soit 0 (ne s'applique pas), 1 (tout à fait en désaccord), 2 (plutôt en désaccord), 3 (plutôt d'accord) et 4 (tout à fait d'accord). Les réponses ne s'appliquant pas (0) sont considérées ici comme des données manquantes et sont exclues des analyses.

Échantillons

Pour les besoins de la présente étude, deux échantillons de données ont été constitués. Le premier échantillon, constitué aux trimestres d'automne 2007, d'hiver et d'été 2008 à partir de la version *papier* du questionnaire, est composé de 20 245 étudiantes et étudiants répartis dans 868 classes. Le second échantillon, constitué aux trimestres d'automne 2010, d'hiver et d'été 2011 au moyen de la version *en ligne*, est composé de 16 432 étudiantes et étu-

dians répartis dans 1 094 classes. Les données proviennent de huit départements et de deux campus de l'Université du Québec à Rimouski, situés à Lévis et à Rimouski.

Procédures d'administration de l'instrument

Modalité papier. La réglementation en vigueur sous la modalité *papier* prévoit que la direction de l'université fasse parvenir le questionnaire au professeur concerné. Au cours des neuf dernières heures de son cours, le professeur libère une heure pour permettre à ses étudiants de le remplir individuellement. Une fois remplis, les questionnaires sont remis à un étudiant désigné par le groupe. Celui-ci les dépose dans une enveloppe prévue à cette fin, la scelle, y appose sa signature et la remet à l'Administration.

Modalité en ligne. Au maximum 24 jours précédant les cinq derniers jours ouvrables du trimestre en cours, les étudiants peuvent remplir le questionnaire. Avant le début de la période d'évaluation, le professeur responsable d'une activité d'enseignement explique la procédure à suivre aux étudiants.

Au début du trimestre suivant, les résultats de l'évaluation sont transmis au professeur et analysés par le comité de gestion du programme auquel l'activité d'enseignement est rattachée à des fins académiques. 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 chacune des dimensions. Un résultat global peut être calculé.

Plan d'analyses

Dimensionnalité. Afin d'évaluer la dimensionnalité de l'instrument analysé, des analyses factorielles exploratoires et confirmatoires sont réalisées sur les 21 items du questionnaire. Les items ayant trait à la dimension *Contexte du Cours* ne sont pas considérés dans ces analyses. Quatre modèles (M_1 , M_5 , M_{51s} , M_{51h}) sont pris en considération dans l'analyse confirmatoire. Ceux-ci permettent de déterminer si la structure du questionnaire est unidimensionnelle (M_1) ou multidimensionnelle (M_5). Dans le cas où elle est multidimensionnelle, les modèles ultérieurs précisent s'il existe un facteur de second ordre (M_{51s}), ou un facteur hiérarchique (M_{51h}). Ces modèles s'apparentent à ceux développés par Brunner et al. (à paraître) pour modéliser l'intelligence.

La démarche est appliquée aux versions *papier* et *en ligne* afin de vérifier si ces dimensions se retrouvent dans les deux modalités. Les paragraphes suivants explicitent les modèles élaborés.

Modèle à un facteur M_1 . Le premier modèle considère que l'ensemble du questionnaire ne mesure qu'un seul et unique facteur. Celui-ci correspond à la *Compétence générale en enseignement* telle que rapportée dans la documentation. Dans ce modèle, la valeur de chaque item s'explique par la présence de ce facteur général et par l'erreur de mesure associée à chacun des items. La faiblesse de ce modèle découle du fait que l'apport des facteurs spécifiques n'est pas considéré. Il est représenté à la figure 1a.

Modèle à cinq facteurs de premier ordre M_5 . Le modèle à cinq facteurs interreliés, appelé M_5 , est celui qui correspond à la structure théorique du questionnaire. La figure 1b illustre ce modèle. Les cinq facteurs correspondent aux dimensions spécifiques du questionnaire. La faiblesse de ce modèle vient du fait qu'il ne tient pas compte de la présence d'un facteur général.

Modèle de second ordre M_{5I_5} . Ce modèle permet de concilier les deux modèles précédents. Il sous-tend l'hypothèse que la *Compétence générale en enseignement* est un facteur de second ordre qui influence les facteurs primaires. La particularité de ce modèle, reproduit à la figure 1c, est que les items ne sont pas en relation directe avec ce facteur de second ordre ; ce sont plutôt les facteurs primaires qui le sont.

Modèle hiérarchique M_{5I_h} . Ce modèle, illustré à la figure 1d, est dit hiérarchique, bi-factoriel ou à facteurs nichés, selon la terminologie employée (Brunner et al., à paraître). La particularité de celui-ci est que chaque item est expliqué à la fois par la présence du facteur hiérarchique, celle d'un facteur primaire, et par l'erreur de mesure associée aux items. Tel qu'il a été mentionné antérieurement, ce facteur hiérarchique s'interprète comme étant la *Compétence générale en enseignement*.

Estimation des paramètres et processus de sélection de modèles. L'estimation des paramètres de chacun des modèles est faite en utilisant l'analyse factorielle confirmatoire (AFC ; Brown, 2006). L'AFC permet de considérer des structurations des facteurs de premier et de second ordres, ainsi que hiérarchiques. L'algorithme du maximum de vraisemblance robuste est utilisé. Le modèle sélectionné est celui qui maximise les indices d'adéquation et dont les paramètres sont interprétables (Brown, 2006 ; Harvey, 2009, 2011 ; Hélie, 2006). Les indices d'adéquation sont basés sur la vraisemblance d'un modèle. Les indices utilisés dans le cadre de la présente étude sont le chi-carré (χ^2), l'erreur moyenne quadratique d'approximation (RMSEA) et l'indice général d'adéquation (GFI). Les valeurs critiques sont $RMSEA \leq 0,06$, $GFI \geq 0,90$, et le χ^2 ne doit pas être significatif. Ce dernier indice est cependant vulnérable

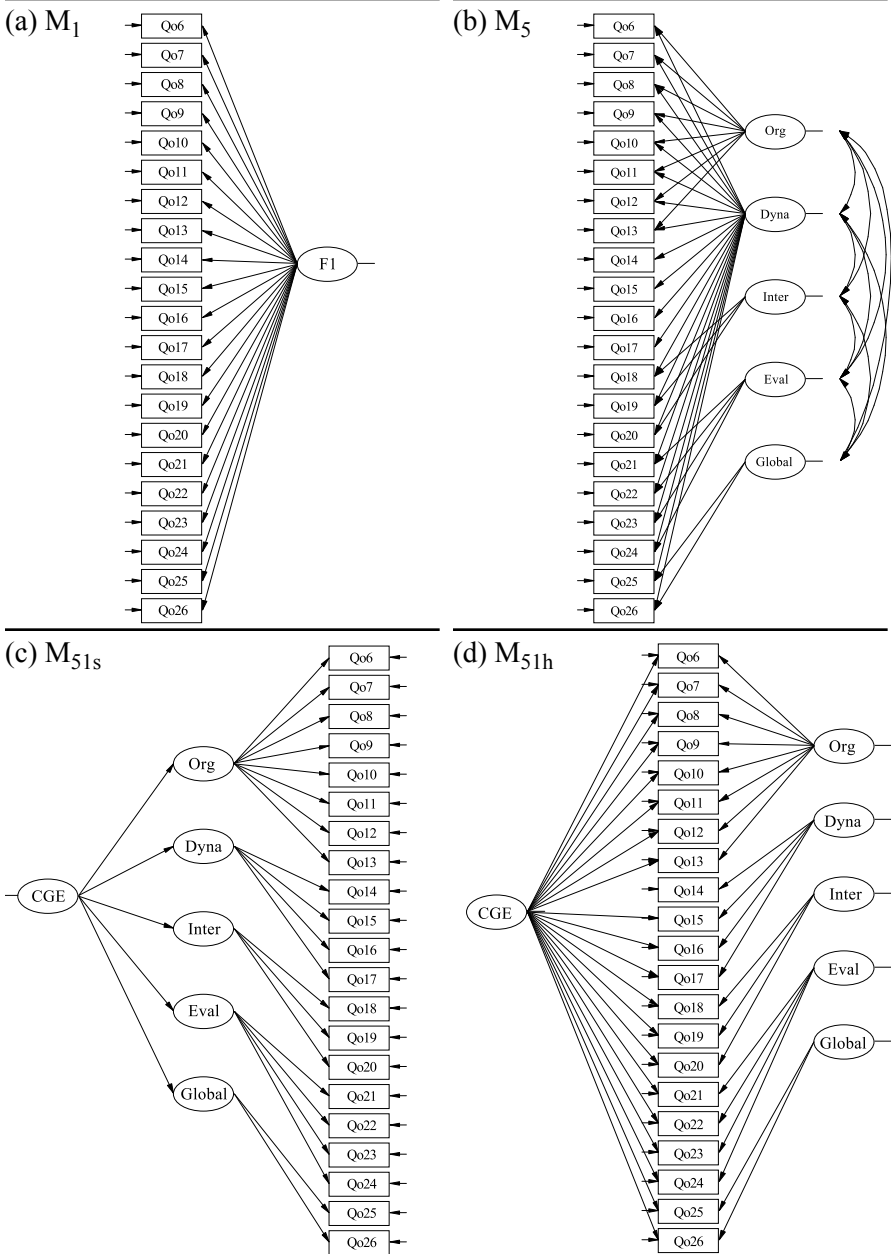


Figure 1 **Représentation graphique des modèles (a) à 1 facteur (M_1), (b) à 5 facteurs (M_5), (c) de second ordre avec 5 facteurs au premier niveau et un au second niveau (M_{51s}), (d) avec la Compétence générale en enseignement (CGE) comme facteur hiérarchique (M_{51h}).**

Org. = Organisation, Dyna. = Dynamisme, Eval. = Évaluation, Global = Appréciation générale
 Inter. = Interaction

aux échantillons de grande taille. Par conséquent, l'adéquation du modèle doit être interprétée à la lumière de l'ensemble des indices. Les analyses sont réalisées à partir de la procédure *Calis* du progiciel statistique SAS 9.2.

Fiabilité et autres analyses. Les indices de fiabilité calculés sont l'alpha de Cronbach et l'*oméga*. Ils sont calculés pour chaque dimension et pour l'ensemble du questionnaire. Les taux de réponses sont calculés pour chacune des modalités *en ligne* et *papier*. Un test du χ^2 est utilisé afin de tester la différence entre les modalités. Les taux de satisfaction de l'enseignement sont calculés par item, par dimension et pour l'ensemble du questionnaire. Des tests *t* sont calculés pour détecter la présence de différences significatives entre les modalités.

Résultats

Hypothèse 1 : taux de réponses

Le tableau 2 présente trois types d'information : les nombres de répondants potentiels et réels et les taux de réponses pour les modalités *papier* et *en ligne*. L'hypothèse d'un écart significatif entre leurs taux de réponses respectifs trouve écho au sein des données. On remarque en effet une forte différence entre les taux de réponses, laquelle se chiffre à 26,3 %. La modalité *papier* a engendré une plus grande participation chez les étudiants que la modalité *en ligne*.

Tableau 2
Taux de réponses selon la modalité d'évaluation

Type	Nombre de répondants potentiels	Nombre de répondants réels	Taux de réponses (%)	Dif. ^{a,b}
<i>Papier</i>	23 834	20 245	84,9	26,3***
<i>En ligne</i>	28 055	16 432	58,6	

^a Un test du χ^2 a été calculé pour tester la différence entre les modalités papier et en ligne.

^b *** $p < 0,001$

Hypothèse 2 : taux de satisfaction

Ensemble du questionnaire. Les taux de satisfaction des enseignements pour les années scolaires 2007-2008 (*papier*) et 2010-2011 (*en ligne*) sont quasi similaires et témoignent de la satisfaction générale face aux enseignements reçus. Ils s'établissent respectivement à 3,61 sur 4 (ÉT: 0,46) et 3,52 sur 4 (ÉT: 0,52), soit 90,3 % et 88,0 % de satisfaction. En retirant des analyses les items nichés sous la dimension *Contexte du cours*, les taux de satisfaction varient très peu ; ils se chiffrent à 3,59 (89,8 %) pour l'année 2007-2008 (ÉT: 0,52) et à 3,49 (87,3 %) pour l'année 2010-2011 (ÉT: 0,59). La figure 2 montre que les écarts se situent sur le plan des activités d'enseignement qui obtiennent des taux de réponses inférieurs à 30 %. En particulier, pour les activités où les taux de réponses se situent entre 10 % et 30 %, le taux de satisfaction pour la modalité en ligne est inférieur de 10 % par rapport à la moyenne des activités. Il est de 3,13 sur 4, soit 78,3 %. Cette situation touche environ 4 % des activités (46 activités sur 1 094). Pour les activités ayant des taux de réponses se situant entre 30 % et 100 %, aucune différence n'est observée entre les modalités *en ligne* et *papier*.

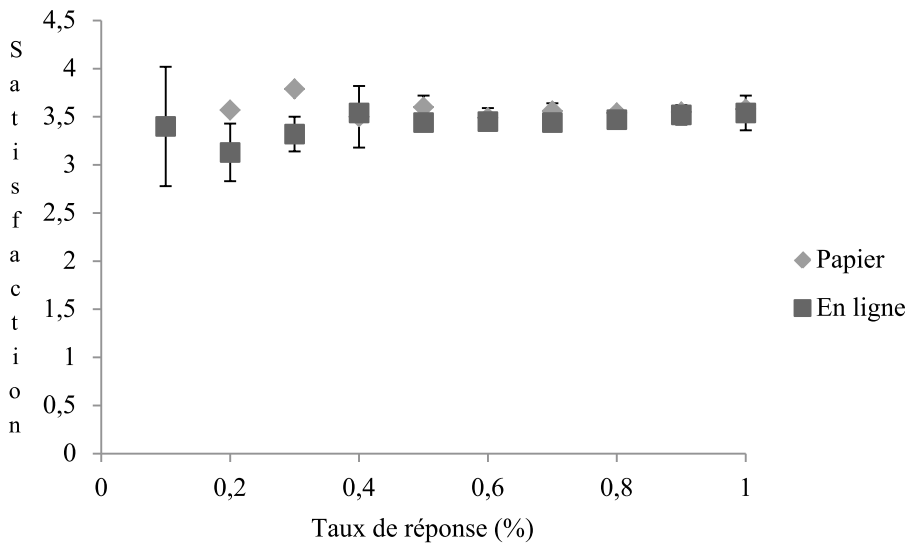


Figure 2. *Taux de satisfaction en fonction du taux de réponses pour les versions en ligne et papier.* (Les barres d'erreur indiquent les intervalles de confiance des moyennes $p \leq 0,01$).

Par item. Le tableau 3 reprend, pour les deux modalités d'intérêt, les moyennes et les écarts-types calculés pour les 26 items du questionnaire. On constate dans l'ensemble un bon degré de satisfaction qui oscille entre 3,77 sur 4 (94,3% : «Le professeur ou le chargé de cours manifeste de l'intérêt pour son enseignement») et 3,26 sur 4 (81,5% : «L'enseignement reçu stimule mon intérêt à l'étude»). Sur les 26 comparaisons de moyenne effectuées, la totalité donne lieu à des différences significatives du point de vue statistique, à la faveur de la modalité *papier*. Des différences variant de 0,06 à 0,17 sont enregistrées.

Par dimension. Lorsqu'il s'agit de considérer les différences entre les modalités *en ligne* et *papier* pour chacune des six dimensions du questionnaire, le portrait demeure le même. À la lecture du tableau 4, on constate que les taux de satisfaction pour la modalité *papier* sont supérieurs à ceux observés pour la modalité *en ligne*, et ce, quelle que soit la dimension considérée. Ici encore, considérant l'effet de grandeur, les différences sont jugées négligeables.

Tableau 3
Moyenne et écart-type de chacune des questions selon la modalité d'évaluation

Items Libellés	Papier		En ligne		Dif. ^{ab}
	M	ET	M	ET	
1 L'horaire du cours me convient bien.	3,69	0,67	3,59	0,73	0,10***
2 Les notes de cours et, s'il y a lieu, les livres obligatoires sont rendus disponibles au bon moment.	3,68	0,66	3,61	0,69	0,07***
3 Le groupe d'étudiants (nombre, homogénéité/hétérogénéité, formation préalable, etc.) favorise mes apprentissages.	3,71	0,59	3,63	0,63	0,07***
4 La salle de cours présente des conditions favorables aux apprentissages (éclairage, chauffage, ventilation, espace, etc.).	3,60	0,72	3,44	0,80	0,17***
5 Les équipements dans la salle de cours présentent des conditions favorables aux apprentissages (appareils audiovisuels, informatiques, etc.).	3,76	0,53	3,70	0,55	0,06***
6 Le plan de cours est clair.	3,68	0,61	3,59	0,69	0,09***
7 Dans l'ensemble, le plan de cours est respecté.	3,67	0,64	3,58	0,70	0,09***
8 Le cours est bien structuré.	3,49	0,77	3,40	0,83	0,09***
9 La matière vue dans le cours respecte la description du cours.	3,64	0,64	3,57	0,70	0,08***
10 Les moyens pédagogiques utilisés (lectures, activités, exercices, etc.) me permettent d'acquérir des habiletés, attitudes et connaissances.	3,45	0,78	3,36	0,85	0,09***
11 La charge totale de travail de ce cours est normale.	3,52	0,74	3,44	0,76	0,08***
12 Le professeur ou le chargé de cours manifeste une disponibilité qui m'encourage à le consulter au besoin.	3,62	0,66	3,52	0,74	0,10***
13 Les cours du professeur ou du chargé de cours sont bien préparés.	3,62	0,68	3,52	0,76	0,10***
14 Le professeur ou le chargé de cours manifeste de l'intérêt pour son enseignement.	3,77	0,54	3,67	0,62	0,09***
15 L'enseignement reçu stimule mon intérêt à l'étude.	3,40	0,84	3,26	0,93	0,15***
16 Dans l'ensemble, les formules pédagogiques utilisées m'aident à atteindre les objectifs du cours.	3,47	0,76	3,35	0,83	0,11***
17 Au moment opportun, le professeur ou le chargé de cours sait faire ressortir les points importants de la matière.	3,55	0,72	3,43	0,80	0,12***
18 Le professeur ou le chargé de cours répond clairement aux questions des étudiants.	3,57	0,71	3,46	0,80	0,11***
19 Les explications du professeur ou du chargé de cours me permettent de faire des liens entre les notions théoriques et les applications pratiques.	3,56	0,71	3,45	0,79	0,11***
20 L'encadrement offert dans ce cours favorise mon apprentissage.	3,54	0,71	3,42	0,79	0,13***
21 Les moyens d'évaluation utilisés me permettent de vérifier l'atteinte des objectifs du cours.	3,50	0,75	3,38	0,81	0,12***
22 Les consignes pour la rédaction des travaux sont claires.	3,46	0,77	3,35	0,84	0,11***
23 Le professeur ou le chargé de cours remet les résultats de mes évaluations dans un délai raisonnable.	3,68	0,62	3,57	0,71	0,10***
24 Mon travail est évalué et commenté d'une manière juste et constructive.	3,56	0,72	3,47	0,77	0,09***
25 Le contenu de ce cours m'a fait progresser dans mon champ d'études.	3,45	0,81	3,35	0,86	0,10***
26 Ce cours m'a permis d'intégrer des connaissances, habiletés et attitudes.	3,53	0,75	3,42	0,80	0,10***

^a Des tests *t* ont été calculés pour tester la différence entre les modalités.

^b ****p* < 0,001

Tableau 4
*Moyenne et écart-type de chacune des dimensions
selon la modalité d'évaluation*

Dimensions	Nombre d'items	Papier		En ligne		Dif. ^{ab}
		M	ET	M	ET	
Contexte du cours	5	3,69	0,39	3,60	0,44	0,09***
Organisation et clarté	8	3,60	0,51	3,51	0,58	0,09***
Dynamisme, intérêt et habiletés d'enseignement	4	3,55	0,61	3,43	0,69	0,12***
Interaction avec les étudiants et les étudiantes	3	3,56	0,64	3,45	0,72	0,11***
Évaluation et rétroaction	4	3,56	0,56	3,46	0,62	0,10***
Appréciation générale	2	3,49	0,75	3,39	0,81	0,10***

^a Des tests t ont été calculés pour tester la différence entre les modalités.

^b *** $p < 0,001$

Hypothèse 3 : fiabilité (alpha de Cronbach)

Qu'il soit informatisé ou non, le questionnaire dans son entier montre des mesures de fiabilité plus qu'acceptables lorsque celle-ci est estimée par l'alpha de Cronbach (*papier*: 0,95; *en ligne*: 0,96). Le tableau 5 met en évidence les fiabilités estimées pour les six dimensions du questionnaire. À l'exception de la dimension *Contexte du cours*, plutôt singulière, l'alpha de Cronbach calculé pour chacune des dimensions se situe au-delà du seuil habituellement souhaité (0,80). Comme attendu, l'écart observé entre les fiabilités des modalités *papier* et *en ligne* pour le questionnaire dans son entier et pour l'ensemble des dimensions est très faible: il oscille entre 0,01 (*Interaction avec les étudiants et étudiantes*, *Appréciation générale*) et 0,05 (*Contexte du cours*). En reprenant les calculs tout en omettant les items sous-jacents à la dimension *Contexte du cours*, la mesure de fiabilité croît pour chacune des modalités (*papier*: 0,96; *en ligne*: 0,97).

Tableau 5
Fiabilité (alpha de Cronbach et Oméga) des différentes dimensions selon la modalité d'évaluation

Dimensions	Nombre d'items	Alpha		Oméga	
		<i>Papier</i>	<i>En ligne</i>	<i>Papier</i>	<i>En ligne</i>
Contexte du cours	5	0,60	0,65	–	–
Organisation et clarté	8	0,89	0,91	0,10	0,10
Dynamisme, intérêt et habiletés d'enseignement	4	0,87	0,89	0,01	0,04
Interaction avec les étudiants et les étudiantes	3	0,89	0,90	0,06	0,04
Évaluation et rétroaction	4	0,81	0,83	0,15	0,14
Appréciation générale	2	0,93	0,94	0,32	0,29
Compétence générale en enseignement	21	0,96	0,97	0,97	0,95

Hypothèse 4 : Dimensionnalité (version en ligne)

Analyse exploratoire. Dans un premier temps, une analyse factorielle exploratoire a été réalisée sur les 21 items du questionnaire. Cette analyse permet de préciser le nombre de facteurs orthogonaux (non corrélés) présents dans les données. Aucune rotation des facteurs n'a été réalisée afin de maintenir intacte la contribution de la composante principale. L'analyse révèle la présence d'un facteur principal qui explique 61 % (racine latente $\lambda = 12,81$) de la variance totale. Tous les items saturent dans ce facteur avec des pondérations qui se situent dans l'intervalle [0,57, 0,86]. Le second facteur n'explique que 5 % de la variance (racine latente $\lambda = 1,05$).

Analyses confirmatoires. Le tableau 6 présente le résultat des analyses confirmatoires. Le modèle M_1 à une dimension doit être rejeté sur la base des indices d'adéquation. Le modèle à cinq dimensions M_5 est acceptable, ce qui suggère que le modèle sur lequel le questionnaire est construit est valide. Par contre, les corrélations entre les cinq facteurs retenus par le questionnaire sont très élevées. Elles se situent entre 0,81 et 0,97. Les cinq facteurs possèdent une validité discriminante faible et sont largement redondants. Le modèle M_{5Is} qui considère qu'il existe un facteur de second ordre est également acceptable, mais doit être rejeté par souci de parcimonie. Il possède sensiblement les mêmes indices d'adéquation que le modèle M_5 , mais sa structure de second ordre est plus complexe d'interprétation. Ces deux modèles ne sont pas ceux qui décrivent le mieux les données.

Tableau 6
Indices d'adéquation obtenus pour les modèles étudiés (en ligne)

<i>Modèles</i>	<i>Paramètres</i>	<i>GFI</i>	<i>RMSEA</i>	χ^2	<i>dl</i>	<i>p</i>
M ₁	42	0,81	0,102	28664	189	0,0001
M ₅	52	0,90	0,076	15170	179	0,0001
M _{51s}	52	0,89	0,077	15726	179	0,0001
M _{51h}	63	0,93	0,062	9576	170	0,0001

Le modèle hiérarchique M_{51h} est celui qui décrit le mieux l'ensemble des données. Il possède l'indice GFI le plus élevé et le RMSEA le plus bas. Le χ^2 demeure élevé et est significatif, ce qui est attribuable à la grande taille de l'échantillon. Ce modèle suggère qu'une *Compétence générale en enseignement* (CGE) influence d'une manière positive le résultat de chacun des items. En effet, tous les items sont de bons indicateurs de cette compétence. Le tableau 7 présente les saturations standardisées associées à ce facteur. Elles varient entre 0,51 et 0,88.

Les saturations des items associés aux facteurs primaires sont beaucoup plus faibles, bien que leurs apports soient significatifs. Ainsi, sept items (10, 11, 12, 14, 16, 17 et 20) ont des saturations inférieures ou égales à 0,12. Pour cinq autres (13, 18, 19, 21 et 22), elles se situent entre 0,20 et 0,30. Pour les autres items, des saturations supérieures à 0,30 sont enregistrées.

La présence d'une dimension hiérarchique a un impact considérable sur la fiabilité des dimensions primaires. Les coefficients de fiabilité *Oméga* (tableau 5) sont inférieurs à 0,30 pour les dimensions *Organisation et clarté*, *Dynamisme*, *intérêt et habiletés d'enseignement*, *Interaction avec les étudiants et étudiantes*, *Évaluation et rétroaction* et *Appréciation générale*. En contrepartie, le coefficient de fiabilité du facteur hiérarchique CGE est excellent (0,95).

Tableau 7
Pondérateur, erreur et variance expliquée (R²) pour chacun des facteurs primaires et la Compétence Générale en Enseignement (CGE) (en ligne)

Items	Org.	Dyna.	Inter.	Éval.	Global	CGE	Erreurs	R ²
Q6	0,45					0,64	0,62	0,61
Q7	0,56					0,63	0,54	0,71
Q8	0,36					0,78	0,51	0,74
Q9	0,36					0,75	0,55	0,70
Q10	0,04					0,82	0,57	0,68
Q11	0,04					0,55	0,83	0,31
Q12	0,04					0,70	0,72	0,49
Q13	0,23					0,79	0,57	0,68
Q14		0,06				0,74	0,67	0,54
Q15		0,52				0,86	0,00	1,00
Q16		0,07				0,88	0,47	0,78
Q17		0,00				0,81	0,59	0,65
Q18			0,23			0,80	0,55	0,70
Q19			0,21			0,85	0,48	0,77
Q20			0,12			0,88	0,46	0,79
Q21				0,20		0,79	0,58	0,66
Q22				0,24		0,69	0,68	0,54
Q23				0,36		0,51	0,78	0,39
Q24				0,46		0,70	0,55	0,70
Q25					0,54	0,82	0,18	0,97
Q26					0,36	0,83	0,42	0,83

Org. = Organisation, Dyna. = Dynamisme, Inter. = Interaction, Éval. = Évaluation,
 Global = Appréciation générale

Hypothèse 4 : Dimensionnalité (version papier)

Analyse exploratoire. L'analyse factorielle exploratoire révèle des résultats similaires à la version *en ligne*. Un facteur principal explique 57% ($\lambda = 11,96$) de la variance totale et le second facteur ($\lambda = 1,06$) n'en explique que 5%. Tous les items saturent dans ce facteur principal avec des pondérations qui se situent entre 0,58 et 0,85.

Analyses confirmatoires. Ici encore, le modèle hiérarchique M_{51h} est celui qui décrit le mieux l'ensemble des données. Il possède l'indice GFI le plus élevé et le RMSEA le plus bas. Le tableau 8 présente les indices d'adéquation pour chacun des modèles. Le tableau 9, quant à lui, montre les saturations standardisées associées à chacun des facteurs du modèle choisi M_{51h} . Il apparaît clair que la structure factorielle de la version *papier* est similaire à celle de la version *en ligne*. Elle confirme la présence d'un facteur *CGE* et, à un degré moindre, la présence des autres facteurs primaires.

Comme c'est le cas pour la version *en ligne*, les coefficients de fiabilité *Oméga* des dimensions spécifiques sont faibles (tableau 5). Ils sont inférieurs à 0,35 pour les dimensions *Organisation et clarté*, *Dynamisme*, *intérêt et habiletés d'enseignement*, *Interaction avec les étudiants et étudiantes*, *Évaluation et rétroaction* et *Appréciation générale*. Le coefficient de fiabilité de la dimension hiérarchique *CGE* est excellent (0,97).

Tableau 8
Indices d'adéquation obtenus pour les modèles étudiés (papier)

<i>Modèles</i>	<i>Paramètres</i>	<i>GFI</i>	<i>RMSEA</i>	χ^2	<i>dl</i>	<i>p</i>
M_1	42	0,83	0,096	31339	189	0,0001
M_5	52	0,91	0,070	15877	179	0,0001
M_{51s}	52	0,91	0,072	16454	179	0,0001
M_{51h}	63	0,94	0,059	10652	170	0,0001

Tableau 9
Pondérateur, erreur et variance expliquée (R²) pour chacun des facteurs primaires et la Compétence générale en enseignement (CGE) (papier)

<i>Items</i>	<i>Org.</i>	<i>Dyna.</i>	<i>Inter.</i>	<i>Éval.</i>	<i>Global</i>	<i>CGE</i>	<i>Erreurs</i>	<i>R²</i>
Q6	0,44					0,59	0,68	0,54
Q7	0,53					0,61	0,59	0,65
Q8	0,37					0,76	0,53	0,71
Q9	0,34					0,73	0,59	0,65
Q10	0,03					0,79	0,61	0,62
Q11	0,01					0,53	0,85	0,28
Q12	0,04					0,64	0,77	0,41
Q13	0,23					0,79	0,57	0,68
Q14		0,21				0,70	0,68	0,54
Q15		0,15				0,83	0,53	0,71
Q16		0,00				0,85	0,53	0,72
Q17		0,00				0,79	0,62	0,62
Q18			0,23			0,78	0,58	0,66
Q19			0,30			0,83	0,48	0,78
Q20			0,15			0,85	0,51	0,74
Q21				0,17		0,77	0,61	0,62
Q22				0,24		0,63	0,74	0,45
Q23				0,39		0,53	0,75	0,44
Q24				0,46		0,66	0,59	0,64
Q25					0,46	0,79	0,41	0,83
Q26					0,49	0,81	0,32	0,89

Org. = Organisation, Dyna. = Dynamisme, Inter. = Interaction, Éval. = Évaluation,
 Global = Appréciation générale

Discussion

Le premier objectif de cet article était de déterminer les qualités psychométriques de l'instrument utilisé à l'Université du Québec à Rimouski pour l'évaluation de l'enseignement. Le second objectif consistait à vérifier l'impact de la modalité de passation (*papier vs en ligne*) sur les qualités psychométriques de l'instrument. Ces deux objectifs se déploient en quatre hypothèses.

En lien avec le premier objectif, il est possible d'affirmer que le questionnaire se structure selon un modèle hiérarchique (hypothèse H₄). Ce modèle établit une structure où deux ensembles de facteurs ont une influence simultanée afin d'expliquer chacun des items. Dans cette perspective, les items du questionnaire ont une double fonction. Tout d'abord, ils sont d'excellents indicateurs d'un facteur hiérarchique que la documentation interprète comme étant une *Compétence générale en enseignement (CGE)*. Cette interprétation est justifiée dans le présent contexte car l'outil utilisé se compose principalement d'items associés à l'acte d'enseignement. La fiabilité *Oméga* de ce facteur est forte et explique de 95 % à 97 % de la variance totale. L'analyse hiérarchique suggère cependant que les items, parce qu'ils saturent peu, sont des indicateurs faibles de la présence de facteurs spécifiques liés à l'acte d'enseigner. La fiabilité de ces facteurs spécifiques est faible lorsque le coefficient *Oméga* est considéré. Ces résultats suggèrent que l'instrument est pleinement valide lorsqu'il est interprété globalement comme un indicateur de la qualité de l'enseignement. Les décisions académiques ou liées à la carrière professorale basées sur le résultat des dimensions spécifiques sont cependant peu fiables. Ces résultats corroborent la thèse d'Abrami (Abrami et al., 2007 ; d'Apollonia & Abrami, 1997) et s'insèrent dans la perspective que certains construits en psychologie et en éducation possèdent une structure hiérarchique (Brunner et al., à paraître).

Ainsi, Abrami recommande que des items généraux (par ex., *Le contenu de ce cours m'a fait progresser dans mon champ d'études*) soient utilisés à des fins sommatives d'évaluation et que ceux liés à des dimensions plus spécifiques (par ex., *L'enseignement reçu stimule mon intérêt à l'étude*), dont la fiabilité est plus faible, soient réservés à des fins formatives. Sur le plan méthodologique, ces résultats montrent les limites des analyses basées sur des modèles de premier ordre qui ne considèrent pas la présence d'un facteur hiérarchique. Des analyses qui considèrent plusieurs modèles alternatifs incluant des modèles de premier et second ordres ainsi que des modèles hiérarchiques permettent de mettre en opposition différentes interprétations et de vérifier celles qui sont valides et celles qui ne le sont pas. Notre analyse suggère qu'un modèle hiérarchique décrit mieux les données que les modèles alternatifs. Ce modèle met clairement en évidence le double rôle des items, ce que ne fait pas un modèle de premier ordre.

En lien avec le second objectif, les résultats de la présente recherche confirment que les modalités de passation ont une influence sur le taux de réponses, ce qui corrobore l'hypothèse H₁. On remarque dans certains cas une

influence de la modalité sur le taux de satisfaction des enseignements, ce qui va à l'encontre de l'hypothèse H_2 . Finalement, tel qu'attendu, la modalité n'influence pas la fiabilité (hypothèse H_3) ni la dimensionnalité des évaluations (hypothèse H_4).

L'influence des modalités sur le taux de réponses est importante. De 85 % en version *papier*, il est inférieur à 60 % en version *en ligne*, soit une perte d'environ 25 % du taux de réponses. Ceci est conforme à ce que rapporte la documentation. Quant à l'influence sur le taux de satisfaction, elle est faible, voire même négligeable considérant la grandeur de l'effet. Elle est globalement de 2% en faveur de la version *papier*. Cet écart pourrait avoir différentes origines. Il pourrait être attribuable à la plus grande sévérité des étudiantes et des étudiants en 2010-2011 qu'en 2007-2008, occasionnée par exemple par le renouvellement du personnel, ou à des changements dans la programmation de l'université. Cette hypothèse est cependant rejetée, car la baisse ne se manifeste que lorsque le taux de réponses est inférieur à 30%. À partir de ce seuil, les évaluations sont 10% inférieures à la moyenne de l'ensemble, ce qui affecte environ 4% des activités d'enseignement. Si la baisse était attribuable à d'autres facteurs que la modalité de passation, elle affecterait l'ensemble des taux de réponses ; ce qui n'est pas le cas ici. Dans le cas présent, la baisse du taux de réponses entraîne, selon toute vraisemblance, un phénomène statistique où la transformation en pourcentage à partir d'un nombre faible de répondants rend variable l'estimation. Dans ce contexte, la présente étude suggère qu'il faut être prudent lorsqu'un taux de réponses inférieur à 30% est enregistré, et ce, tant dans une perspective sommative que formative. La documentation suggère que des taux de réponses supérieurs à 70% devraient être atteints afin de considérer comme valide une évaluation. En ce sens, notre étude ne corrobore pas ce taux.

Quant à la fiabilité et à la dimensionnalité du questionnaire, elles ne sont pas affectées par les modes de passation. Une structure hiérarchique dont les coefficients de fiabilité sont similaires est observée dans les deux cas.

Conclusion

En conclusion, l'évaluation de la qualité de l'enseignement par les étudiantes et les étudiants peut être considérée comme un processus valide et fiable lorsque l'interprétation qui est faite à partir de l'instrument de mesure s'insère dans une perspective où le résultat global est considéré. La présente étude ne permet pas de valider les interprétations basées sur une décomposition de l'outil en facteurs spécifiques. Au mieux, ces interprétations peuvent être soulevées dans un cadre formatif afin de susciter la réflexion, mais ne devraient pas être faites dans une perspective sommative liée à la carrière professorale. Par ailleurs, l'utilisation d'une version *en ligne* n'invalide pas le processus. Il faut cependant être conservateur lorsque des taux de réponses inférieurs à 30 % sont enregistrés. Les taux de satisfaction sont alors 10 % inférieurs à la moyenne des évaluations.

Sur le plan méthodologique, la distinction d'un facteur hiérarchique qui s'interprète en termes de compétence générale en enseignement suggère que les études de validation doivent considérer la présence de facteurs de second ordre dans leurs analyses. Ce facteur a été observé à partir de l'analyse des réponses obtenues pour deux années académiques et pour des cours offerts en classe régulière. Dans ce mode d'enseignement, la compétence en enseignement de la ressource professorale est particulièrement sollicitée, ce qui explique la dominance de cette composante hiérarchique. Dans d'autres modalités d'enseignement, telles que des cours offerts à distance, en tutorat et en laboratoire, cette composante générale en enseignement peut être moins importante que lors d'un enseignement régulier. La conclusion qu'une compétence générale en enseignement influence les évaluations ne peut donc pas s'appliquer, à l'heure actuelle, à d'autres formes d'enseignement que l'enseignement en classe régulière. L'hypothèse que l'importance de la composante hiérarchique varie avec les modalités du cours a déjà été soulevée dans la documentation.

Par ailleurs, selon certaines hypothèses, l'enthousiasme de la ressource professorale, et le recours à des évaluations favorables des travaux et examens des étudiants et des étudiantes (« *evaluation leniency* ») seraient des facteurs qui biaisent l'EEE et invalident l'interprétation en termes de compétence générale en enseignement. Une analyse approfondie des questions relatives à l'existence de biais potentiels dépasse l'objectif de cet article. Néanmoins, des auteurs tels que Marsh et Dunkin (1992) considèrent l'EEE comme l'une des

meilleures façons d'évaluer l'enseignement offert dans le cadre d'un cours par une ressource professorale et qu'elle est peu affectée par des biais potentiels.

Finalement, il existe la possibilité qu'il y ait plus qu'un facteur hiérarchique. Marsh et Dunkin (1992) ont relevé jusqu'à quatre facteurs hiérarchiques associés au SEEQ. Cependant, les facteurs supplémentaires sont généralement corrélés et se présentent telle une décomposition de la compétence générale en enseignement en plusieurs facteurs de second ordre. Pour cette raison, la présente étude s'est limitée à rechercher un seul facteur hiérarchique.

NOTE

1. L'utilisation du terme *professeur* désigne dans le présent texte toutes les ressources professorales qui offrent une activité d'enseignement créditée à l'ordre d'enseignement universitaire.

RÉFÉRENCES

- Abrami, P. C., d'Apollonia, S., & Rosenfield, S. (2007). The dimensionality of student ratings of instruction: What we know and what we do not. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 385-445). Berlin: Springer.
- Abrami, P. C., Rosenfield, S., & Dedic, H. (2007). The dimensionality of student ratings of instruction: An update on what we know and what we do not. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 446-456). Berlin: Springer.
- American Educational Research Association (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.
- Brunner, M., Nagy, G., & Wilhelm, O. (à paraître). A tutorial on hierarchically structured constructs. *Journal of Personality*.
- Byrne, M., & Flood, B. (2003). Assessing the teaching quality of accounting programmes : an evaluation of the Course Experience Questionnaire. *Assessment & Evaluation in Higher Education*, 28(2), 135-145. doi:10.1080/02602930301668
- Cates, W. M. (1993). A small-scale comparison of the equivalence of paper-and-pencil and computerized versions of student end-of-course evaluations. *Computers in Human Behavior*, 9, 401-409. doi:10.1016/0747-5632(93)90031-M

- Cohen, E. (2005). Student evaluations of course and teacher: factor analysis and SSA approaches. *Assessment & Evaluation in Higher Education*, 30(2), 123-136. doi:10.1080/0260293042000264235
- Crews, T. B., & Curtis, D. F. (2011). Online course evaluations: Faculty perspective and strategies for improved response rates. *Assessment & Evaluation in Higher Education*, 36(7), 865-878. doi:10.1080/02602938.2010.493970
- 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
- De Ketele, J. M., & Gérard, F. M. (2005). La validation des épreuves d'évaluation selon l'approche par compétences. *Mesure et évaluation en éducation*, 28(3), 1-26.
- Dommeyer, C. J., Baum, P., Hanna, R. W., & Chapman, K. S. (2004). Gathering faculty teaching evaluations by in-class and on-line surveys: their effects on response rates and evaluations. *Assessment & Evaluation in Higher Education*, 29(5), 611-623. doi:10.1080/02602930410001689171
- Donovan, J., Mader, C. E., & Shinsky, J. (2006). Constructive student feedback: On-line vs. traditional course evaluations. *Journal of Interactive Online Learning*, 5(3), 283-296.
- Donovan, J., Mader, C. E., & Shinsky, J. (2007). Online vs. traditional course evaluation formats: Students perceptions. *Journal of Interactive Online Learning*, 6(3), 158-180.
- Feldman, K. A. (1976). The superior college teacher from the students' view. *Research in Higher Education*, 5, 243-288. doi:10.1007/BF00991967
- Fontaine, S. (2009). Des expériences actuelles d'évaluation des enseignements vers des démarches adaptées aux 2^e et 3^e cycles. In M. Romainville & C. Coggi (Eds.), *L'évaluation de l'enseignement par les étudiants : Approches critiques et pratiques innovantes* (pp. 123-144). Bruxelles : De Boeck.
- Grim Fidelman, C. (2007). *Course evaluation surveys: in-class paper surveys versus voluntary online surveys* (Unpublished doctoral dissertation). Boston College University, Boston.
- Harvey, L. (2009). L'échafaudage lors de la supervision en milieu professionnel : études des modalités et un modèle. *Mesure et évaluation en éducation*, 32(1), 55-83.
- Harvey, L. (2011). Hidden Markov models and learning in authentic situations. *Tutorials in Quantitative Methods for Psychology*, 7(2), 12-21.
- Hélie, S. (2006). An introduction to model selection: Tools and algorithms. *Tutorials in Quantitative Methods for Psychology*, 2(1), 1-10.
- Hess, M., Barron, A. E., Carey, L., Hilbelink, A., Hogarty, K., Kromrey, J. D., Phan, H., & Schullo, S. (2005). From the learners' eyes: Student evaluation of online instruction. *Proceedings of the National Educational Computing Conference*, 26, 1-23.
- Laveault, D., & Grégoire, J. (2002). *Introduction aux théories des tests en psychologie et en sciences de l'éducation* (2^e éd.). Bruxelles : De Boeck.
- Layne, B. H., DeCristoforo, J. H., & McGinty, D. (1999). Electronic versus traditional student ratings of instruction. *Research in Higher Education*, 40(2), 221-232. doi:10.1023/A:1018738731032
- Liegle, J., & McDonald, D. S. (2004). Lessons learned from on-line vs. paper-based computer information students' evaluation system. In D. Colton, J. Tastle, M. Hensel, & A. Amjad Addullat (Eds.), *The proceedings of the 21st annual information systems education conference*, ISECON (pp. 1-12). Newport: AITP Foundation for Information Technology Education.

- Louis, R., Jutras, F., & Hensler, H. (1996). Des objectifs aux compétences, l'évaluation de la formation initiale des maîtres. *Revue canadienne de l'éducation*, 21(4), 414-432. doi:10.2307/1494894
- Linn, R. L., & Miller, M. D. (2005). *Measurement and assessment in teaching* (9^e éd.). Columbus: Prentice Hall.
- Marsh, H. W. (1983). Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics. *Journal of Educational Psychology*, 75(1), 150-166. doi:10.1037//0022-0663.75.1.150
- Marsh, H. W. (1984). Students' evaluation of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76(5), 707-764. doi:10.1037//0022-0663.76.5.707
- 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
- Marsh, H. W., & Dunkin, M. J. (1992). Students' evaluations of university teaching: A multidimensional perspective. In J. Smart (Ed.), *Higher education: Handbook of theory and research* (vol. 8, pp. 143-233). New York: Agathon.
- Monsen, S., Woo, W., Mahan, C., & Miller, G. (2005). *Online course evaluations: Lessons learned*. Presentation at the CALI Conference for Law School Computing, Chicago, Illinois.
- Poissant, H. (1996). *L'évaluation de l'enseignement universitaire*. Paris: Logiques.
- 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
- Reid, I. C. (2001). Reflections on using Internet for the evaluation of course delivery. *Internet and Higher Education*, 4, 61-75. doi:10.1016/S1096-7516(01)00048-3
- Romainville, M. (2009). Une expérience collective de critères de qualité. In M. Romainville & C. Coggi (éds), *L'évaluation de l'enseignement par les étudiants: Approches critiques et pratiques innovantes* (pp. 145-166). Bruxelles: De Boeck.
- Romainville, M., & Coggi, C. (éds) (2009). *L'évaluation de l'enseignement par les étudiants: Approches critiques et pratiques innovantes*. Bruxelles: De Boeck.
- Thorpe, S. W. (2002, June). *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.
- Toland, M. D., & De Ayala, R. J. (2005). A multilevel factor analysis of students' evaluation of teaching. *Educational & Psychological Measurement*, 65(2), 272-296. doi:10.1177/0013164404268667
- Université du Québec à Rimouski (UQAR) (2011a). *Expérimentation de l'évaluation de l'enseignement en ligne (2009-2011). Rapport final – Déposé à la Commission des études, 13 septembre 2011*. Rimouski: Bureau du Doyen des études de premier cycle.
- Université du Québec à Rimouski (UQAR) (2011b). *Résolution CE-470-5637 de la Commission des Études, 13 septembre 2011*. Rimouski: Université du Québec à Rimouski.
- Wong, A., & Fitzsimmons, J. (2008). Student evaluation of faculty: An analysis of survey results. *U21Global Working Paper Series*, 3, 1-7.

- Younès, N. (2009). L'évaluation de l'enseignement par les étudiants comme seuil de changement. In M. Romainville & C. Coggi (Eds.), *L'évaluation de l'enseignement par les étudiants : Approches critiques et pratiques innovantes* (pp. 191-210). Bruxelles : De Boeck.
- Zaltman, G., Pinson, C. R. A., & Angelmar, R. (1973). *Metatheory and consumer research*. New York, NY: Holt, Reinhart and Winston.

Date de réception : 28 juin 2012

Date de réception de la version finale : 20 septembre 2012

Date d'acceptation : 26 septembre 2012