

Mesure et évaluation en éducation



Les étudiants sont-ils compétents pour évaluer la qualité des enseignements ?

Pascal Detroz et Jean-Guy Blais

Volume 35, numéro 3, 2012

Date de réception : 31 mai 2012

Date de réception de la version finale : 8 octobre 2012

Date d'acceptation : 11 octobre 2012

URI : <https://id.erudit.org/iderudit/1024668ar>

DOI : <https://doi.org/10.7202/1024668ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (imprimé)

2368-2000 (numérique)

[Découvrir la revue](#)

Citer cet article

Detroz, P. & Blais, J.-G. (2012). Les étudiants sont-ils compétents pour évaluer la qualité des enseignements ? *Mesure et évaluation en éducation*, 35(3), 5–30.
<https://doi.org/10.7202/1024668ar>

Résumé de l'article

Cet article répond à cette question centrale : « Les étudiants sont-ils compétents pour évaluer la qualité des enseignements ? » en l'abordant selon deux angles distincts. Le premier a trait aux attitudes et représentations des différentes parties prenantes quant à cette question. Ces attitudes ne sont pas univoques bien qu'elles soient, en tendance, favorables à l'implication des étudiants dans les dispositifs d'évaluation des enseignements. Le second a trait à la validité de l'information ainsi récoltée. La plupart des recherches et les méta-analyses effectuées soutiennent plutôt la validité de l'EEE. Ce n'est cependant pas toujours le cas et cette validité est au moins partiellement en débat. Les auteurs plaident, dans les conclusions, pour que l'écologie locale et le contexte, notamment disciplinaire, soient pris en compte lors d'analyses de la validité qui devraient, selon eux, envisager des méthodologies plus naturalistes.

Les étudiants sont-ils compétents pour évaluer la qualité des enseignements ?

Pascal Detroz

Université de Liège

Jean-Guy Blais

Université de Montréal

MOTS CLÉS : évaluation des enseignants par les étudiants, attitude, validité, fidélité, approche naturaliste

Cet article répond à cette question centrale : « Les étudiants sont-ils compétents pour évaluer la qualité des enseignements ? » en l'abordant selon deux angles distincts. Le premier a trait aux attitudes et représentations des différentes parties prenantes quant à cette question. Ces attitudes ne sont pas univoques bien qu'elles soient, en tendance, favorables à l'implication des étudiants dans les dispositifs d'évaluation des enseignements. Le second a trait à la validité de l'information ainsi récoltée. La plupart des recherches et les méta-analyses effectuées soutiennent plutôt la validité de l'EEE. Ce n'est cependant pas toujours le cas et cette validité est au moins partiellement en débat. Les auteurs plaident, dans les conclusions, pour que l'écologie locale et le contexte, notamment disciplinaire, soient pris en compte lors d'analyses de la validité qui devraient, selon eux, envisager des méthodologies plus naturalistes.

KEY WORDS: student ratings, attitude, validity, reliability, naturalistic approach

In this paper we answered a central question: "Do students have enough skills to assess teaching's quality?" through two different indicators. The first one related to stakeholders various attitudes and representations about this question, knowing that they (attitudes and representations) change from one stakeholder to one other but are, as a trend, in favor of including students within the teaching quality assessment. The second one deals with the concept of validity of collected information. Most of the studies and meta-analysis support the validity of the process even though some validity might still be discussed. We plead, as a conclusion, to also take into account local ecology and context (mainly disciplinary) in validity analysis which may lead, according to us, to a more naturalistic methodology.

PALAVRAS-CHAVE: *avaliação do ensino pelos estudantes, atitude, validade, fidelidade, abordagem naturalista*

Este artigo responde, segundo dois ângulos distintos, a esta questão central: “Os estudantes são competentes para avaliar a qualidade do ensino?” O primeiro tem a ver com as atitudes e as representações das diferentes partes sobre esta questão. Estas atitudes não são unívocas, mesmo que elas sejam, tendencialmente, favoráveis à implicação dos estudantes nos dispositivos de avaliação do ensino. O segundo tem a ver com a validade da informação recolhida desta forma. A maior parte das investigações e meta-análises efetuadas sustentam a validade da avaliação do ensino pelos estudantes, não obstante ser posta causa e, pelo menos, parcialmente discutida. Os autores defendem, nas suas conclusões, que a ecologia local e o contexto, nomeadamente disciplinar, sejam tomados em conta nas análises da validade, na senda, segundo eles, de metodologias mais naturalistas.

Note des auteurs – Toute correspondance peut être adressée comme suit : Pascal Detroz, Pédagogie de l’enseignement supérieur, Université de Liège, bât. B3 IFRES-SMART, Traverse des Architectes 4, 4000 Liège 1, Belgique, téléphone +32 4 3662281, télécopieur +32 4 36622700, ou par courriel à l’adresse suivante [p.detroz@ulg.ac.be].

Introduction

Appréhender la question « les étudiants sont-ils compétents pour évaluer la qualité des enseignements ? » avec, comme référent théorique, l'approche par compétences aurait eu du sens. Cependant, la notion de compétence est loin d'être transparente et de renvoyer à un sens unifié (Rey, Carette, Defrance, & Kahn, 2006). De nombreux auteurs se sont penchés sur ce concept durant ces 20 dernières années. C'est notamment le cas des travaux de Perrenoud (1997), Roegiers (2001), Beckers (2002), Le Boterf (2005) ou encore Tardif (2006) et chacun de ces auteurs propose une définition différente du concept de compétence. Ces différences s'expliquent essentiellement par le contexte dans lequel ceux-ci centrent leurs recherches. Ainsi, selon Prégent, Bernard, et Kozanitis (2009), ils sont influencés par une série de variables exogènes qui sont leur réalité culturelle (européenne ou nord-américaine), leur discipline (didactique, psychologie, docimologie) ou par leur milieu de travail (scolaire ou professionnel). Même si leurs définitions sont différentes, ces auteurs s'accordent *a minima* sur le fait que l'approche par compétences cherche à développer la possibilité par les apprenants de mobiliser un ensemble de ressources, internes ou externes, afin de résoudre une situation problème appartenant à une famille de situations. Dès lors, l'évaluation des compétences consiste à vérifier la mobilisation de certaines ressources dans le contexte ainsi défini.

À la lumière de l'approche par compétences, la question générique de cet article deviendrait : « les étudiants parviennent-ils à mobiliser un ensemble de ressources, internes ou externes, afin de résoudre une situation problème (donner son avis sur la qualité d'un cours) appartenant à une famille de situations (l'évaluation des cours) ? » Cette question serait légitime si et seulement si l'on acceptait comme vraies un certain nombre d'assertions telles que

- 1) les étudiants mobilisent des ressources internes et externes pour remplir les questionnaires d'évaluation et
- 2) donner son avis sur un enseignement est une situation problème.

Or, si l'on considère les caractéristiques d'une situation problème selon De Vecchi et Carmona-Magnaldi (2002), celle-ci doit avoir du sens ; être liée à un obstacle repéré défini comme surmontable ; faire naître un questionnement chez les élèves ; créer une ou des ruptures amenant à déconstruire le ou les modèles explicatifs initiaux s'ils sont inadaptés ou erronés ; correspondre à

une situation complexe, si possible liée au réel, pouvant ouvrir sur différentes réponses acceptables et différentes stratégies utilisables ; déboucher sur un savoir d'ordre général et faire l'objet d'un ou de plusieurs moments de métacognition. Il semble dès lors difficile de considérer que répondre à un questionnaire d'évaluation de cours soit une situation problème. L'action de remplir de tels formulaires ne peut donc être l'expression d'une compétence si l'on considère l'acception théorique de ce dernier terme.

Par ailleurs, si l'on envisage l'acception commune de cette notion de compétence, telle celle livrée par le dictionnaire Robert de 2006, la définissant comme étant « une connaissance approfondie, reconnue, qui confère le droit de juger ou de décider en certaine matière », la question « les étudiants sont-ils compétents pour évaluer la qualité des enseignements ? » est en résonnance avec deux champs de la documentation en évaluation des enseignements par les étudiants (EEE). Le premier est lié aux perceptions, représentations ou attitudes des différentes parties prenantes (enseignants, étudiants, autorités académiques) quant à la capacité des étudiants à porter un jugement sur les cours qu'ils ont suivis. Le second s'intéresse à la validité des données issues de l'EEE. Si celles-ci étaient valides, elles refléteraient d'une certaine manière une image adéquate de la qualité de l'enseignement. De ce fait, les étudiants pourraient être considérés comme des observateurs pertinents (compétents ?) dont l'avis a une valeur heuristique.

Les perceptions, représentations et attitudes des parties prenantes

Les étudiants sont-ils compétents pour évaluer les enseignements dont ils sont les bénéficiaires ? Les docimologues pourraient être tentés de répondre : ni plus ni moins que les enseignants quand ils évaluent la qualité des apprentissages de leurs étudiants. Et si cette dernière pratique est relativement peu remise en question malgré quelques alertes fondamentales de la docimologie critique, dans la lignée des travaux de Pieron (1963), il n'en va pas de même quand le « bic rouge change de main »¹. En effet, l'évaluation des enseignements par les étudiants souffre encore de nombreuses résistances observables dans le discours de certains enseignants universitaires.

« L'évaluation des enseignants bride la liberté académique », « derrière l'évaluation de l'enseignement se cache l'évaluation de l'enseignant », « dans l'évaluation des enseignements par les étudiants, seuls les commentaires écrits

sont porteurs de sens», «les matières difficiles bénéficient d'une évaluation défavorable», «il suffirait de diminuer ses exigences vis-à-vis des étudiants pour obtenir une évaluation plus avantageuse», «il ne faut pas rendre les évaluations obligatoires pour les étudiants : seuls ceux qui ont des choses pertinentes à dire et qui veulent les dire doivent être interrogés», «il faudrait compiler les différentes mesures pour avoir un seul score de qualité d'un enseignement donné», «les enseignements adressés aux petits groupes bénéficient d'évaluations plus favorables», «lors de l'évaluation des enseignements par les étudiants, il faut interroger les étudiants qui ont été souvent absents du cours : ils représentent la frange des moins satisfaits», «lors de l'évaluation des enseignements par les étudiants, il ne faut interroger que les étudiants ayant été présents à l'ensemble des cours ; seuls ceux-ci peuvent porter un jugement sur la qualité du processus qu'ils ont vécu», «le résultat des évaluations n'est pas pris en compte dans le cadre des promotions», etc.

Il n'aura pas échappé au lecteur que certains de ces propos sont antinomiques. L'évaluation des enseignements est-elle donc condamnée à vivre et à assumer ces paradoxes ? Peut-être pas. Il est possible que la plupart des contradictions ainsi relevées ne soient qu'apparentes et résultent d'un cadre conceptuel trop frustré et le plus souvent laissé dans l'ombre.

Sous le vocable générique d'«évaluation des enseignements» se cachent de multiples situations, allant de l'enseignant demandant à la fin du cours à l'un ou l'autre étudiant s'il a compris un point de matière précis, jusqu'aux évaluations externes dans lesquelles un groupe d'experts se penche sur un programme d'études. Si ces deux situations poursuivent des objectifs proches – récolter de l'information pour réguler un dispositif d'enseignement –, il semble évident qu'elles ne partagent pas nécessairement les mêmes logiques, ni les mêmes méthodes. Cette diversité est le miroir de la richesse des processus d'enseignement. Il serait contreproductif de la limiter, par exemple, en standardisant les processus d'évaluation. *A contrario*, il semble important de préciser et comprendre le cadre conceptuel dans lequel une évaluation des enseignements particulière prend place. D'une part, parce que cela peut éviter un certain nombre de phantasmes liés à une procédure qui, même si elle vise l'enseignement, n'épargne pas complètement l'enseignant en tant que personne et, d'autre part, parce que c'est un jalon nécessaire à la cohérence du dispositif évaluatif mis en place.

Il est probable que la plupart des critiques émanant des enseignants résultent du déficit cumulé lié à l'incompréhension du cadre conceptuel (résultant souvent d'un manque de clarification de celui-ci par les autorités) et à la méconnaissance de la documentation traitant de l'évaluation des enseignements². Ce n'est pas pour autant qu'elles sont injustifiées. Les auteurs y reviendront dans leurs conclusions.

L'évaluation des enseignements par les étudiants reste donc chez certaines parties prenantes un sujet polémique. Comme le soulignent Bernard, Postiaux, et Salcin (2000), il n'est pas rare d'entendre des enseignants remettre en question la valeur et l'efficacité de ce type d'évaluation. Cela est même devenu une antienne chez certains d'être eux.

L'avis des enseignants

Divers auteurs ont cherché à objectiver quantitativement ce sentiment de l'inefficacité de l'EEE. Peut-être les critiques à ce sujet sont-elles exprimées de manière plus virulente par les enseignants en opposition au dispositif? Peut-être existe-t-il une majorité silencieuse moins acerbe et plus en phase avec l'EEE?

Cette question a été traitée en divers endroits de la planète. Si Marsh (1987), Gervais, Nadeau, et Barnabé (1996), Bernard, Postiaux, et Salcin (2000) ainsi que Beran et Rokosh (2009) ont recueilli les attitudes ou les perceptions des enseignants concernant l'EEE en Amérique du Nord, Younès (2006), Detroz (2010), Nasser et Fresko (2002), Chang (2003), Sall (2009), et El Hassan (2009) en ont fait de même respectivement en France, en Belgique, en Israël, en Taiwan, au Sénégal et au Liban.

Ces études donnent des résultats difficilement comparables, faute de clarification théorique. Les divers articles traitent de sujets proches mais en utilisant un vocabulaire qui recouvre, à l'analyse, des réalités différentes. Ainsi, les notions d'évaluation des enseignements, d'évaluation de la qualité des enseignements ou encore d'évaluation des enseignants semblent utilisées indifféremment sans pour autant que les auteurs n'explicitent clairement leur position théorique. De la même manière, lorsque ces auteurs considèrent la capacité des étudiants à fournir une information pertinente sur les enseignements, ils parlent indifféremment de compétences, d'aptitudes ou de connaissances des étudiants.

Ces études montrent toutefois que les enseignants n'ont pas une attitude univoque lorsqu'il s'agit de considérer l'intérêt de l'EEE. Ils sont aussi par-tagés sur le fait que les étudiants aient les compétences (ou aptitudes, connaissances) pour évaluer les enseignements. Outre ce constat, il est intéressant d'essayer de percevoir comment peut s'expliquer cette divergence de point de vue.

C'est ce que Kogan, Schoenfeld-Tacher, et Hellyer (2010) ont tenté de faire en catégorisant les attitudes des enseignants en fonction de leur genre, le type de contrat qui les lie à l'université et leur rang académique. Cette enquête, portant sur 75 réponses d'enseignants provenant du Colorado, montre que seul le genre a un effet sur les attitudes face à l'EEE. Les enseignantes exprimeraient plus d'émotions négatives face à une évaluation critique et seraient moins enclines à en utiliser les résultats à des fins de régulation (ces conclusions sont en phase avec l'étude de Nasser et Fresko (2002)). Ces résultats démontrent toutefois l'utilité, non plus seulement de récolter les attitudes des enseignants, mais aussi de catégoriser ces attitudes en fonction de variables personnelles ou contextuelles.

L'avis des étudiants

Peu d'études abordent les croyances des étudiants au sujet de l'EEE. Pourtant, si les étudiants doutent du dispositif, ils ne seront pas dans un état d'esprit idéal pour y contribuer. Une recherche de la documentation a fourni cinq documents exploitables pour approcher les attitudes des étudiants à ce sujet.

Les travaux de Thivierge et Bernard (1996) montrent que les étudiants montréalais interrogés sont confiants quant à l'utilité du dispositif EEE et revendiquent ce droit à l'évaluation.

L'étude de Heine et Maddox (2009) établit que les étudiants ont le sentiment d'évaluer les enseignements avec sérieux, même si, contrairement à leurs homologues québécois, ils doutent un peu de l'efficacité globale du dispositif.

Deux autres études sur le sujet (Balam & Shannon, 2010; Sojka, Gupta, & Deeter-Schmelz, 2002) comparent les attitudes des enseignants et celles des étudiants sur une série de croyances liées à l'EEE. Dans les deux études, les publics interrogés sont relativement sévères avec le dispositif d'EEE, même si les étudiants sont en général plus cléments face à celui-ci.

La dernière étude est celle de Younès (2007) qui compare les attitudes des étudiants, des enseignants et des autorités vis-à-vis de l'EEE. Elle précise que ces trois populations affirment de concert que l'EEE est importante, qu'elle devrait être régulière et qu'elle serait à encourager. Cependant, au-delà de ces positions de principes communes entre ces acteurs, elle relève des désaccords en ce qui concerne les modalités pratiques de mise en œuvre, par exemple l'institutionnalisation de la procédure d'EEE (partie prenante du programme ou non), les sources d'information utilisées (les étudiants uniquement ou l'usage de sources multiples), ou la diffusion des résultats.

L'avis des autorités académiques

Le soutien des autorités académiques est sans conteste un levier puissant pour l'acceptation et l'utilisation des résultats issus de l'EEE. De nombreux auteurs, dont Paulsen et Feldman (1995) ainsi que Lucas (1990), ont insisté sur l'importance d'une culture universitaire qui soutienne l'enseignement et son évaluation. Paradoxalement, toutefois, il semble que même si un nombre croissant d'universités ont recours à l'EEE (Seldin, 1993), les autorités académiques ont des perceptions ambiguës au sujet de ces méthodes d'évaluation. Dans l'enquête de Wright et O'Neil (1995), les autorités académiques – issues d'universités ancrées dans différents pays (Canada, Europe, États-Unis, Australie) – devaient classer 36 stratégies permettant d'améliorer la qualité de l'enseignement. Sur ces 36 stratégies, la récolte de données auprès des étudiants en milieu d'année se classe en 21^e position. Quand la récolte se fait à la fin de l'année, cette procédure de recueil d'information glisse à la 34^e position. Ce résultat montre que l'utilité de la démarche d'EEE n'est pas toujours perçue par les autorités, alors que leur soutien à ce dispositif est un facteur déterminant de son efficacité. Il semble donc utile et nécessaire de mieux comprendre les perceptions et opinions des décideurs sur ce thème. Or, peu de recherches ont été menées à ce sujet. Outre celles précédemment citées de Younès (2007), relevons celles de Desjardins et Bernard (2002) et de Morgan, Sneed, et Swinney (2003) dont la conclusion principale est que les autorités académiques ont des attitudes ambivalentes, mais globalement soutenantes, à l'égard de l'EEE.

Desjardins et Bernard (2002) proposent une étude qualitative axée sur l'interview semi-structurée de 10 administrateurs. Leurs conclusions sont que, pour ces décideurs, l'évaluation des enseignements joue un rôle important pour réguler la qualité des programmes mais qu'il est plus facile, à leur niveau, de mettre en place des stratégies d'amélioration des enseignements indépendantes des évaluations.

L'étude de Morgan et ses collègues (2003), quant à elle, était centrée sur une enquête envoyée aux administrateurs d'universités américaines. Dans cette étude, 57,8 % des autorités académiques pensent que l'évaluation des enseignements par les étudiants est toujours, ou la plupart du temps, une indication pertinente de la qualité d'un enseignement. La principale conclusion de leur étude est que, par rapport aux enseignants, les autorités académiques sont plus enclines à penser que l'évaluation de l'enseignement par les étudiants mesure réellement la qualité de celui-ci.

Dans l'étude de Younès (2007), le principal fait marquant est que les décideurs appellent à utiliser les résultats à des fins de contrôle.

Les conclusions que l'on peut tirer de l'étude des perceptions, représentations et attitudes des parties prenantes quant à la pertinence de l'information issue du dispositif d'EEE (et de la *compétence* dont font preuve les étudiants pour fournir cette information) sont difficilement interprétables. Ces perceptions sont, la plupart du temps, bienveillantes par rapport au dispositif d'évaluation et les étudiants sont jugés assez souvent comme des évaluateurs fiables. Toutefois, on observe à travers toutes ces études – quels que soient les méthodes, contextes géographiques, temporalités et acteurs interrogés – un nombre relativement important de résistances face à cette idée. Il serait intéressant de mener des études statistiques ou qualitatives pour comprendre les raisons de ces doutes. Sont-ils issus d'un profil bien particulier, par exemple celui d'enseignants réfractaires, par principe, à l'évaluation ? Ou alors d'enseignants allergiques à l'enseignement ? Ou d'enseignants issus d'une épistémologie particulière ? Les dispositifs d'EEE bouleversent-ils à ce point des éléments profondément ancrés comme la liberté académique ou la place de l'étudiant dans la relation didactique ?

Qu'en est-il par ailleurs de variables contextuelles ? Le type de dispositif – par exemple formatif vs sommatif – de l'EEE est-il de nature influencer sur ces attitudes ? Quels sont les mécanismes médiateurs entre les dispositifs, les profils des enseignants et leurs attitudes au sujet de l'EEE ? Des études complémentaires mériteraient d'être menées pour y voir plus clair.

La validité de l'EEE

Une autre manière de s'interroger sur les compétences des étudiants à porter un regard évaluatif sur les enseignements est de vérifier si le jugement ainsi effectué reflète réellement les enseignements dont ils ont bénéficié. Cela

revient à vérifier que les données issues du dispositif d'EEE permettent de faire des inférences correctes, par exemple lorsqu'il s'agit de sélectionner des enseignants sur la base de leurs performances pédagogiques (visée certificative). Si ce n'était pas le cas, les fondements mêmes de l'EEE seraient remis en question.

Lorsqu'un docimologue s'intéresse à la qualité d'un processus d'évaluation, il se préoccupe en premier lieu de la validité de celui-ci. À ce stade, il semble important de clarifier le propos. Si l'on se réfère à la définition de l'évaluation de Stufflebeam (1980), celle-ci est définie comme un recueil de données permettant un jugement menant lui-même à une régulation. Les sources potentielles d'invalidité sont donc multiples. Premièrement, elles peuvent résulter d'un recueil de données biaisées ; soit parce que l'instrument de récolte de l'information (ici, le questionnaire proposé aux étudiants) est mal conçu, soit parce que la population y répondant n'est pas représentative (par exemple, lorsque le taux de réponse est très faible).

Ces données vont ensuite être traitées et transformées en mesures. Il s'agit d'une seconde source de biais. En effet, si le modèle qui permet cette transformation est inopportun, même si les données de base sont fiables, la mesure sera biaisée. Ce sera le cas, par exemple, si les données sont regroupées de manière inadéquate.

Enfin, même si les mesures sont de qualité, il faut s'assurer que les inférences qui s'appuient sur elles soient justifiées. C'est une troisième source de biais. À ce sujet, Abrami (2001) note que les résultats sont souvent utilisés de manière abusive par les commissions de nomination des enseignants qui ne comprennent pas de manière exhaustive les données qui leur sont proposées.

À noter que la définition du concept de la validité proposée, en 1999, par un collectif de structures majeures – l'*American Educational Research Association* (EARA), l'*American Psychological Association* (APA) et le *National Council on Measurement in Education* – soutient, à tout le moins partiellement, cette vision de la validité. Cette définition précise que l'analyse de la validité est un processus dont le but est d'accumuler des preuves démontrant le caractère approprié des inférences qui sont faites à partir des résultats d'un test donné. La validité a trait, dès lors, au degré de conviction que l'on peut avoir, à la suite de l'analyse de ces preuves, quant à l'exactitude des interprétations qui sont faites à partir du test et quant au caractère approprié de la façon dont celles-ci sont réalisées.

Un des premiers éléments marquants de cette définition est que la validité ne s'applique pas à un test décontextualisé mais bien en usage et que l'interprétation des résultats d'un test se fait en contexte. En conséquence :

- C'est bien l'utilisation d'un test et l'interprétation des résultats qui en sont issus qui sont l'objet de la validation, pas le test en lui-même. Dans ce cadre, et malgré l'usage courant, dire qu'un test est valide sans en préciser le contexte de validation n'a pas de sens.
- Si c'est l'utilisation et l'interprétation qui sont l'objet de l'étude de validité, le type et la quantité de preuves nécessaires varient en fonction du contexte. Les critères à appliquer à un test formatif ne sont pas nécessairement identiques à ceux que l'on applique à un test certificatif.
- Si le concept de validité est un concept qui apparaît comme étant relativement simple, l'étude de la validité est cependant quelque chose de nécessairement complexe qui prend en compte l'écologie du milieu dans lequel le test a été mis en œuvre.

Traditionnellement, il sera nécessaire de combiner plusieurs éléments d'analyse pour mener une étude de la validité. Selon Sireci (2009), cinq types de preuves doivent être recueillies : celles portant sur le contenu, sur le processus, sur le critère, sur la fidélité, sur la conséquence et sur la praticabilité. Dans le présent article, les auteurs présenteront ces preuves en lien avec l'EEE.

La preuve axée sur le **contenu** se réfère à la manière dont les items d'un questionnaire EEE reflètent réellement les éléments importants en lien avec la qualité de l'enseignement (et seulement ceux-là). En d'autres termes, ces items doivent former un échantillon représentatif de tous les items auxquels on aurait théoriquement pu faire appel pour l'évaluation.

La preuve axée sur le **construit** exprime le fait que les données récoltées à l'aide d'un dispositif EEE sont réellement, et de manière adéquate, en rapport avec l'entité sous-jacente que l'on cherche à étudier : l'enseignement. Ainsi, il faut s'assurer que le « construit » théorique auquel on s'intéresse soit convenablement opérationnalisé par l'EEE.

La preuve basée sur le **critère externe** s'intéresse à la relation entre les données issues de l'EEE et d'autres données reconnues comme étant en lien avec le même objet d'étude, à savoir l'enseignement. Par exemple, pour vérifier cet aspect de la validité d'un dispositif d'évaluation de l'enseignement

par les étudiants, on observera les liens éventuels entre les données issues du dispositif et d'autres données issues d'autres méthodes, comme l'avis de pairs, l'analyse du portfolio de l'enseignant, etc.

La preuve basée sur l'analyse de la **fidélité** d'un test peut se faire de quatre manières différentes. Il s'agit de :

- La fidélité test-retest : pour être fidèle, un questionnaire d'EEE administré à deux reprises, à deux moments différents, doit produire le même score.
- La fidélité entre formes parallèles : on compare deux questionnaires d'EEE créés à partir du même contenu (par exemple, en séparant aléatoirement les items d'un même test) et proposés en même temps aux mêmes étudiants. Les deux formes doivent fournir des données similaires en termes de moyenne et d'écart type pour que le questionnaire soit valide.
- La fidélité par cohérence interne : cette forme de fidélité est utilisée pour juger la consistance des résultats à travers l'analyse des items d'un questionnaire d'EEE, par exemple à l'aide de l'analyse du R.bis ou de l'alpha de Cronbach.
- La fidélité interjuges : ce type de fidélité est mesuré lorsque deux correcteurs indépendants évaluent la même production. On s'intéressera alors à la variance entre l'avis des étudiants concernant la qualité perçue d'un même enseignement. Différents indicateurs statistiques, par exemple le coefficient de Kappa, peuvent être appliqués et fournir un indicateur de la fidélité interjuges.

La preuve basée sur la **conséquence** évalue les effets (souhaités et non souhaités) de l'application d'un dispositif d'EEE. Plusieurs méthodologies permettent de récolter ce type de preuve : la mesure d'impact, l'observation du comportement des individus évalués (insatisfaction, déprime, etc.) ou toute autre information récoltée après le test. C'est également dans ce cadre que prennent place les débats sur les éventuels « effets pervers » des dispositifs EEE (diminuer les exigences du cours pour obtenir un meilleur score à l'évaluation par les étudiants, tendre à une uniformité des enseignements pour se conformer aux critères en vigueur dans les EEE, etc.).

Ces différentes sources de preuves offrent un cadre riche pour instruire la validité d'un dispositif d'évaluation et de son utilisation ainsi que pour récolter divers éléments permettant d'instruire cette délicate question : « le niveau de validité de l'EEE est-il suffisant pour les soutenir de manière extensive? ».

La validité est certainement l'objet d'étude principal de la documentation abondante traitant de l'EEE. Des centaines d'articles lui sont consacrés. Si, dans cette documentation fournie, la validité de l'EEE a parfois été un peu remise en cause, de très nombreux auteurs défendent l'idée que cette validité est tout à fait acceptable, voire très bonne. Citons, dans cette perspective, des auteurs comme Feldman (1988), Marsh (1980), Cohen (1981) qui polarisent d'ailleurs une grande partie de la documentation abordant l'EEE.

Remarquons d'emblée que déclarer de manière péremptoire que les dispositifs d'évaluation des enseignements par les étudiants sont valides (comme le soutiennent, avec plus ou moins de précautions oratoires, ces auteurs) est en porte-à-faux par rapport à la définition proposée de la validité. En effet, comme précisé ci-avant, celle-ci **n'est pas** un élément attribué à un type de démarche d'évaluation (centrée sur l'avis des étudiants). Elle ne peut être en lien qu'avec la qualité du questionnaire, la qualité du modèle qui permet de transformer les données en mesure et la qualité du jugement et de l'interprétation (les inférences) qui en sont déduites. Tout au plus peut-on garantir la validité d'un dispositif d'évaluation dans un contexte et avec un usage très précisément défini (questionnaire, jugement, inférence). Déclarer qu'un type de dispositif d'évaluation (l'EEE) est valide sans autre forme de commentaire ou d'analyse constitue donc à tout le moins une généralisation abusive.

Dans le cadre de l'EEE, les affirmations de validité ne portent pas sur des questionnaires mais sur un type de dispositif dont le seul point commun est de recueillir l'avis des étudiants sur la base d'un questionnaire, sachant que les différences sont nombreuses : finalités, questionnaires, dispositifs de recueil d'information, d'interprétations, de jugements, etc. Or, la documentation francophone, par exemple l'ouvrage de Romainville et Coggi (2009), démontre à quel point les écologies locales sont prégnantes sur la mise en œuvre d'un tel dispositif. On sait par ailleurs qu'un certain nombre de variables contextuelles influent sur les dispositifs EEE. Dans l'étude de Nasser et Fresko (2002), les attitudes des enseignants étaient en partie fonction de leur domaine disciplinaire.

On le comprendra, la notion d'EEE recouvre des réalités extrêmement diversifiées et, en corollaire, une qualité qui peut être très variable. À ce titre, il semble qu'il faille être extrêmement prudent quand on donne à certaines représentations des enseignants le statut de mythe, en raison du fait que la documentation ne les soutient pas. On peut se demander si derrière ces différences apparentes entre l'avis de certains enseignants et celui de certains

chercheurs ne se cachent simplement pas des contextes (qualité de questionnaire, type d'application du dispositif, structuration et communication des *feedbacks*, etc.) radicalement différents. Et on peut regretter que la documentation anglo-saxonne passe la plupart du temps sous silence le contexte dans lequel les données ont été recueillies.

Cette critique fondamentale de la documentation étant énoncée, il est toutefois intéressant de définir les dispositifs méthodologiques qui ont permis à ces auteurs d'arriver à ces conclusions et de les mettre en lien avec les différentes preuves à recueillir pour statuer sur la validité d'un dispositif d'EEE.

Les preuves axées sur le contenu et le construit

La qualité de l'enseignement n'est pas un concept clairement défini à partir duquel il est aisé de **définir** des indicateurs communément admis et partagés. Il est difficile d'en analyser les preuves axées sur le contenu. Diverses méthodes existent néanmoins, comme le fait de demander à des experts (par ex., des enseignants) de statuer sur la pertinence de chacun des items d'un questionnaire. Mais cette méthode a des limites. Feldman (1988), après avoir recueilli l'opinion des enseignants et des étudiants sur les critères de qualité qu'ils jugent importants à propos d'un enseignement, constate un an plus tard (1989) que les éléments ainsi mis en exergue n'ont pas tous nécessairement un effet sur l'apprentissage, alors que d'autres, qui n'ont pas été pointés, jouent un rôle clé dans les apprentissages.

Une autre méthode utilisée consiste à inverser la logique. Un certain nombre de recherches ont été élaborées ou métaanalysées pour tenter de repérer un éventuel construit systématique dans les questionnaires EEE. L'idée de base est la suivante : si l'on repère un ensemble de facteurs identiques à travers toutes ces études, il est probable que ceux-ci soient l'expression d'un construit commun qui, à défaut d'avoir une base théorique, pourrait représenter une vision partagée de ce qu'est un enseignement de qualité. Marsh (1987), Abrami, d'Appolonia, et Cohen (1990) ainsi que Centra (1993), notamment, ont mené des travaux de ce type. Bien qu'il existe un certain nombre de dimensions communes transversales à leurs travaux, il ne leur a pas été possible d'isoler un seul ensemble de dimensions, ce qui entraîne à considérer la qualité de l'enseignement comme étant un construit multidimensionnel, sans pour autant participer à la définition de ce concept.

À défaut d'apporter des preuves irréfutables liées au contenu ou au construit de l'EEE, la majorité des études se sont concentrées sur **les biais en rapport avec la qualité de l'information récoltée** à l'aide de cette méthode. Si l'EEE permet de récolter des données pertinentes pour évaluer la qualité de l'enseignement, elle ne doit pas subir l'influence de variables exogènes sans lien avec la qualité de celui-ci.

Les études sur le biais s'intéressent à quatre grands types d'effets. Les effets liés à l'enseignant, aux étudiants, au contexte du cours et à la procédure de recueil d'information.

Les auteurs ont analysé les liens entre une série de variables liées aux **enseignants** et l'opinion des étudiants à propos de l'enseignement. Citons notamment l'âge (par ex., Renaud & Murray, 1996), l'ethnie (par ex., Glascock & Ruggerio, 2006), la personnalité (par ex., Erdle, Murray, & Rushton, 1985), le genre (par ex., Centra & Gaubatz, 2000), l'expérience (par ex., Feldman, 1983), la qualité de chercheur (par ex., Marsh & Hattie, 2002), le statut de l'enseignant (par ex., Schuckman, 1990) et l'expressivité (par ex., Abrami, Leventhal, & Perry, 1982).

Des liens entre les variables liées aux **étudiants** et l'opinion de ces derniers sur les enseignements suivis ont été analysés : l'âge (par ex., Centra, 1993), le genre (par ex., Bernard, Keefauver, Elsworth, & Maylor, 1981), le niveau d'études (par ex., Aleamoni & Thomas, 1980), la motivation des étudiants (par ex., Marsh & Cooper, 1981) ou encore la note qu'ils espèrent obtenir à l'examen final (par ex., Aleamoni, 1999).

Les variables en lien avec le **contexte** qui ont été observées sont les suivantes : la taille de l'auditoire, le jour de la semaine où le cours est donné, l'heure à laquelle le cours est dispensé, l'année dans laquelle le cours est planifié, la matière enseignée, l'ancienneté du cours, la difficulté de celui-ci, son statut obligatoire ou optionnel et sa rigueur académique.

Enfin, certains liens liés aux **modalités d'administration** de l'EEE ont été recherchés. C'est le cas de la période où est mené le recueil des données, de l'anonymat, de la présence de l'enseignant lorsque les étudiants remplissent le questionnaire, du recueil avant ou pendant un examen, des buts avoués du dispositif d'évaluation (certificatif ou formatif) ou du type de discours effectué avant le recueil des données.

À noter que certains de ces biais ont été analysés dans le cadre d'**études de laboratoire**. La plus célèbre est probablement celle du Dr Fox (Naftulin, Ware, & Donnelly, 1973) au sein de laquelle un acteur donnait un cours truffé d'erreurs et de non-sens mais avec énormément d'emphase. Cet acteur a reçu une très bonne évaluation alors que les apprentissages étaient impossibles. Certains y ont vu le fait que l'opinion des étudiants dépendait plus du style de l'enseignant que des apprentissages qu'il favorisait. Cette étude a cependant été durement critiquée dans la documentation pour des manquements méthodologiques (par ex., Mash & Ware, 1982).

Les études sur les biais ont souvent donné lieu à des résultats contrastés. Si l'on prend l'exemple de celles traitant des liens entre la note (espérée par l'étudiant) à l'examen et son opinion de l'enseignement, Aléamoni (1999) recense 24 études qui ne montrent pas de lien, 37 qui montrent un lien positif et deux qui montrent un lien négatif. Ce type de lien n'est toutefois pas nécessairement le signe d'un biais. Il existe en fait trois hypothèses à ce sujet (Marsh, 2001) :

- L'hypothèse de validité : les étudiants qui apprennent plus sont ceux qui obtiennent les plus hautes notes et sont les plus satisfaits. Elle est défendue par Marsh (1983).
- L'hypothèse inflationniste : les enseignants qui donnent les plus hautes notes sont « récompensés » par une opinion favorable des étudiants concernant l'enseignement qu'ils prodiguent (Greenwald & Gillmore, 1997).
- L'hypothèse liée aux caractéristiques de l'étudiant : certaines de ces dernières influent à la fois sur les performances des étudiants et sur l'opinion qu'ils ont des enseignements (Howard & Maxwell, 1982).

Ce type de résultats contrastés est fréquent dans l'analyse des biais et des études de laboratoire. Aujourd'hui, la plupart des auteurs considèrent, au vu des résultats de nombreuses métaanalyses, qu'il y a peu de liens entre les diverses variables observées et l'EEE. Pour eux, lorsqu'une relation est présente, elle ne constitue pas un biais ou alors elle est jugée insignifiante (sauf pour l'une ou l'autre variable comme la matière enseignée, le statut du cours (optionnel ou obligatoire)). Certes, on peut partager en partie cette analyse si l'on observe une vision macro de ces données. Ces résultats, observés sur un plan plus micro, restent cependant suffisamment contrastés pour inciter à la prudence dans les analyses. Si un certain nombre d'éléments plaident pour une validité de contenu et de construit de l'EEE, ceux-ci ne semblent pas de nature à permettre de statuer une fois pour toutes sur le sujet. Il serait légitime-

me de nourrir une certaine confiance quant à ces validités, mais il semble néanmoins nécessaire et important de vérifier si ces conclusions très générales sont applicables dans les écologies locales, par définition très particulières.

Les preuves axées sur les critères externes

De nombreux articles ont traité de cette question via l'approche multi-traits-multiméthodes qui consiste à corrélérer les résultats obtenus à l'EEE avec d'autres données liées à l'enseignement ou à des données obtenues en utilisant d'autres méthodes. On trouve une large gamme de données auxquelles ont été confrontés les résultats de l'EEE. C'est le cas, par exemple, de celles liées :

- à l'autoévaluation de l'enseignant : Feldman (1989) effectue une métaanalyse sur 19 études et rapporte une corrélation de 0,29 entre cette variable et le score à l'EEE ;
- à la note reçue par l'étudiant à l'examen : la métaanalyse de Cohen (1981) portant sur 41 études rapporte une corrélation de 0,43 entre cette variable et le score à l'EEE ;
- à l'évaluation de l'enseignement effectuée par les administrateurs : Kulik et McKeachie (1975) rapportent des corrélations allant de 0,47 à 0,62 entre cette variable et le score à l'EEE ;
- à l'évaluation de l'enseignement effectuée par les pairs sur la base d'observations : Feldman (1989) rapporte une corrélation de 0,55 entre cette variable et le score à l'EEE ;
- à l'évaluation de l'enseignement effectuée par les alumni : Braskamp et Ory (1994) rapportent des corrélations variant de 0,40 à 0,75 entre cette variable et le score à l'EEE ;
- à l'évaluation de l'enseignement effectuée par les alumni (vision longitudinale) : Overall et Marsh (1980) demandent à 1 400 étudiants d'évaluer les enseignements qui leur ont été proposés et refont l'évaluation un an après leur diplôme. La corrélation est de 0,83 entre ces deux prises de données ;
- aux données issues d'observations d'observateurs entraînés : Murray (1983) entraîne 49 observateurs (étudiants) à repérer certaines actions de l'enseignement. Ces observateurs relèvent ces actions chez 54 professeurs d'université. Chaque cours a été passé au crible de 18 à 24 fois par les observateurs. La corrélation entre observateurs est de 0,32, mais elle est de 0,77 pour l'ensemble des observateurs et les données issues de l'EEE ;

- aux commentaires écrits des étudiants : Ory, Bradskamp, et Pieper (1980) rapportent des corrélations de 0,93 sur la base des commentaires des étudiants et de 0,84 sur la base d'interviews effectuées avec ces derniers.

On pourrait dès lors conclure, ce que la plupart des auteurs ont fait, que les preuves axées sur les critères externes sont probantes puisque tous les liens sont positifs et démontrent parfois des corrélations très fortes. Le souci avec ce type d'études porte sur le statut des données servant à la comparaison. Est-on certain que les données issues de l'autoévaluation, la note à l'examen, l'évaluation par les pairs, etc., constituent des données fiables à partir desquelles on peut statuer sur la qualité de l'enseignement ? Jenkins (1946), cité par Sireci (2009), met très tôt en garde les psychométriciens contre la dérive d'utiliser des données elles-mêmes non valides à des fins de comparaison.

Les preuves axées sur la fidélité

Marsh (1987) a analysé l'ensemble des résultats de la recherche à ce sujet. Il précise que quelle que soit la manière dont on la définit, la fidélité de l'EEE est remarquablement constante. Bien sûr, il affine un peu cette analyse globale. Il constate, par exemple, que la fidélité intercorrecteurs augmente avec le nombre de juges, d'étudiants. Cette fidélité est de 0,60 pour cinq étudiants, de 0,74 pour 10 étudiants, de 0,90 pour 25 étudiants et de 0,95 pour 50 étudiants. Par ailleurs, la fidélité par cohérence interne du questionnaire dépend ... du questionnaire utilisé. Elle doit donc être calculée au cas par cas. Aleamoni (1999) relève un ensemble de questionnaires fournissant des données ayant une cohérence interne de 0,90.

Peut-on en conclure, pour autant, que les preuves axées sur la fidélité sont probantes ? Les présomptions existent. Il serait hasardeux toutefois de se reposer sur ces études pour affirmer que les données issues d'un questionnaire mis en œuvre à l'échelon local sont *a priori* fidèles. Dans ce cas, la présomption de fidélité semblerait abusive, surtout si l'on utilise un questionnaire qui n'a fait l'objet d'aucune étude psychométrique.

Les preuves axées sur la conséquence

Très tôt s'est posée la question de l'impact de l'EEE sur l'amélioration des enseignements. Trois grands types de méthodologie ont été appliqués :

- demander aux enseignants s'ils pensent que l'EEE a un effet sur leur enseignement et sur les apprentissages des étudiants : cette question a été traitée dans la première partie de l'article ;

- observer l'effet d'un dispositif d'EEE sur les jugements effectués par les étudiants en fin d'année, à travers des études longitudinales : Gray et Bradenbourg (1985) constatent une augmentation de la satisfaction des étudiants, mesurée à travers les vagues successives d'évaluations des enseignements par les étudiants ;
- utiliser la méthode expérimentale : l'idée est de constituer, selon les cas, un ou plusieurs groupes expérimentaux et un ou plusieurs groupes de contrôle. Dans l'expérience de Marsh, Flainer, et Thomas (1975), les EEE pour l'ensemble des groupes sont recueillies en milieu d'année. Les *feedbacks* sont uniquement distribués au groupe expérimental, laissant le groupe contrôle sans rétroaction. En fin d'année, une nouvelle récolte de l'opinion des étudiants permet de vérifier si le groupe expérimental obtient des variations de score différentes de celles du groupe contrôle.

En 1979, Overall et Marsh peaufinent ce plan expérimental en y ajoutant un groupe expérimental qui reçoit le *feedback* accompagné d'un *débriefing* oral. Une vingtaine d'études ont été menées sur cette base et ont été maintes fois métaanalysées. Cohen, en 1981, rapporte par exemple que le *feedback* seul a un effet modeste mais significatif sur l'opinion des étudiants en fin d'année. Cet effet est accentué lorsque le *feedback* est accompagné par un *débriefing*. Plus récemment, Hampton et Reiser (2004) constatent un effet de l'évaluation accompagnée d'une rétroaction, mais seulement dans le cas d'un des deux cours qui ont été analysés. Enfin, notons les travaux de Penny et Coe (2004) qui mènent une méta-analyse sur 12 études portant sur les *feedbacks* en milieu d'année. Pour eux, l'effet de l'EEE est insignifiant sauf si la rétroaction apporte des conseils, voire un suivi dans l'implémentation d'améliorations pédagogiques.

Plusieurs critiques peuvent être apportées par rapport à cette documentation.

Lorsqu'il s'agit d'évaluer l'amélioration du cours (consécutivement à un *feedback*, à un *feedback* accompagné d'un *débriefing* ou à l'absence de *feedback* en milieu d'année), utiliser les données récoltées par l'EEE en fin d'année pour défendre l'idée d'un effet de l'EEE en milieu d'année (afin de justifier la validité de ce dispositif) semble, pour le moins, tautologique.

Par ailleurs, les *feedbacks* en milieu d'année n'influencent pas la note obtenue par les étudiants à l'examen de fin d'année (Marsh et al., 1975, Penny & Coe, 2004). En d'autres termes, les enseignants ayant obtenu un *feedback*, même s'il est accompagné d'un *débriefing*, n'ont pas pu améliorer de manière significative les apprentissages de leurs étudiants.

Enfin, on le constate, ce sont surtout les discussions successives à l'obtention des *feedbacks* (*débriefing*) qui ont un effet sur l'amélioration de l'enseignement (telle qu'elle est définie par ces auteurs). *In fine*, on serait presque en droit de se demander quelle est la plus-value réelle des résultats de l'EEE. N'est-elle pas simplement un prétexte pour aborder des thèmes liés à l'enseignement ? Le *débriefing* portait-il bien sur le résultat des *feedbacks* ou sur d'autres éléments dont l'enseignant était conscient et souhaitait discuter ? Les recherches n'ont pas été suffisamment descriptives pour répondre à cette question, mais l'idée que le principal effet de l'EEE soit d'ouvrir un espace de discussion sur le thème de l'enseignement dans des universités essentiellement préoccupées par la recherche paraît raisonnable.

Conclusion

En premier lieu, les auteurs ont fait la synthèse des articles portant sur les perceptions et attitudes des différents acteurs de l'EEE (enseignants, étudiants et administrateurs). Les conclusions sont que ces études présentent des résultats relativement proches : la plupart des acteurs sont bienveillants ou neutres au sujet de l'EEE et au sujet de l'implication des étudiants dans les dispositifs d'évaluation des enseignements mais une minorité, surtout à chercher parmi la population des enseignants, y reste farouchement opposée.

Les auteurs ont ensuite présenté les principales études portant sur la validité de l'évaluation des enseignements par les étudiants. Certains contrastes existent. La plupart des recherches et les méta-analyses effectuées soutiennent plutôt la validité de l'EEE. Ce n'est cependant pas toujours le cas et cette validité est, dans ces études, au moins partiellement en débat.

Ces contradictions ne semblent pas pouvoir être balayées d'un simple revers de la main. Ce n'est pas parce que les métadonnées à disposition incitent à penser que **globalement** l'EEE est valide et acceptable qu'un dispositif local est forcément valide et acceptable. Les auteurs y perçoivent même un questionnement fondamental : cherche-t-on les biais au bon endroit ?

Certes, il est confortable d'adosser l'EEE à une documentation consistante défendant la validité globale du dispositif. De telles études doivent d'ailleurs être poursuivies, probablement en les ancrant plus avant dans les méthodes de validation d'outils psychométriques, tels que les modèles de réponse à l'item.

On remarquera toutefois que l'EEE met en jeu des processus complexes et variés, à visées et enjeux multiples, impliquant de nombreux acteurs (ayant leurs propres perceptions et opinions), interférant avec des écologies locales prégnantes et pas complètement indépendants de variables contextuelles telles que le champ disciplinaire ou les modalités d'enseignement des enseignants évalués.

La mécanique est complexe et les grains de sable pouvant s'y glisser sont nombreux. À quoi sert-il, par exemple, de proposer aux étudiants un questionnaire valide si le taux de réponses est de quelques points de pourcentage? À quoi sert-il de donner aux enseignants un *feedback* complet et diagnostique s'ils ne savent pas l'interpréter? En posant ces questions, c'est l'usage qui est fait de l'EEE dans nos universités qui est interrogé. Or, cet usage est spécifique et échappe au moins partiellement aux études classiques portant sur la validité.

C'est la raison pour laquelle les auteurs proposent d'adopter une démarche de validation qui inclut, outre les méthodes psychométriques formalistes précédemment évoquées, des méthodes plus naturalistes qui s'attachent à la compréhension des phénomènes en contexte, qui respectent les écologies locales et qui nécessitent une plus forte implication de l'ensemble des parties prenantes. Cette validation mixte des dispositifs EEE à l'échelle locale semble la mieux à même de permettre des boucles de régulation courtes et efficaces permettant l'amélioration constante du dispositif EEE et de ses usages.

NOTES

1. Pour reprendre le titre de l'article publié dans le périodique *15^e jour du mois* en 2003, consacré au sujet de l'évaluation des enseignements par les étudiants.
2. Sur ce sujet, Franklin et Theall (1989) ont mis en évidence un lien positif entre un manque de connaissance au sujet de l'EEE et des opinions négatives au sujet de ce type d'évaluation et de la valeur des *feedbacks* associés.

RÉFÉRENCES

- Abrami, P. C. (2001). *Improving judgements about teaching effectiveness using teacher ratings forms*. In M. Theall, P. C. Abrami, & L. A. Mets (Eds.), *The student ratings debate: Are they valid? How can we best use them?* (pp. 59-87). San Francisco: Jossey Bass.
- Abrami, P. C., d'Apollonia, S., & Cohen, P. A. (1990). Validity of student ratings of instruction: What we know and what we do not. *Journal of Educational Psychology*, 82, 219-231.
- Abrami, P. C., Leventhal, L., & Perry, R. P. (1982). Educational seduction. *Review of Educational Research*, 52, 446-464.
- Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924-1998. *Journal of Personnel Evaluation in Education*, 13(2), 153-166.
- Aleamoni, L. M., & Thomas, G. S. (1980). Differential relationships of student, instructor, and course characteristics to general and specific items on a course evaluation questionnaire. *Teaching of Psychology*, 7(4), 233-235.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Balam, E., & Shannon, D. (2010). Student ratings of college teaching: A comparison of faculty and their students. *Assessment & Evaluation in Higher Education*, 35(2), 209-221.
- Beckers, J. (2002). *Développer et évaluer des compétences à l'école : vers plus d'efficacité et d'équité*. Bruxelles: Labor.
- Beran, T. N., & Rokosh, J. L. (2009). Instructors' perspectives on the utility of student ratings of instruction. *Instructional Science*, 37(2), 171-184.
- Bernard, H., Postiaux, N., & Salcin, A. (2000). Les paradoxes de l'évaluation de l'enseignement universitaire. *Revue des sciences de l'éducation*, 26(3), 625-650.
- Bernard, M. E., Keefauver, L. W., Elsworth, G., & Maylor, F. D. (1981). Sex role behavior and gender in teachers-student evaluations. *Journal of Educational Psychology*, 73, 681-696.
- Braskamp, L. A., & Ory, J. C. (1994). *Assessing faculty work: Enhancing individual and institutional performance*. San Francisco: Jossey-Bass.
- Centra, J. A. (1993). *Reflexive faculty evaluation effectiveness. enhancing teaching and determining faculty effectiveness*. San Francisco: Jossey-Bass.
- Centra J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluation of teaching? *The Journal of Higher Education*, 70(1), 17-33.
- Chang, T. S. (2003, April). *The results of student ratings: The comparison between paper and online surveys*. Paper presented at the annual meeting of American Educational Research Association, Chicago, IL.
- Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, 51(3), 281-309.

- Desjardins, J., & Bernard, H. (2002). Les administrateurs face à l'évaluation des enseignements. *Revue des sciences de l'éducation*, 28(3), 617-648.
- Detroz, P. (2010). *Le point sur les dispositifs d'évaluation des enseignements par les étudiants: pertinence, utilisation, amélioration* (Thèse de doctorat non publiée). Université de Liège, Liège.
- De Vecchi, G., & Carmona-Magnaldi, N. (2002). *Faire vivre de véritables situations-problèmes*. Paris: Hachette Éducation.
- El Hassan, K. (2009). Investigating substantive and consequential validity of student ratings of instruction. *Higher Education Research & Development Journal*, 28(3), 319-333.
- Erdle, S., Murray, H. G., & Rushton, J. P. (1985). Personality, classroom, behavior, and college teaching effectiveness: A path analysis. *Journal of Educational Psychology*, 77, 394-407.
- Feldman, K. A. (1983). Seniority and experience of college teachers as related to evaluations they receive from their students. *Research in Higher Education*, 18, 3-124.
- Feldman, K. A. (1988). Effective college teaching from the students' and faculty's view: Matched or mismatched priorities? *Research in Higher Education*, 28, 291-344.
- Feldman, K. A. (1989). Instructional effectiveness of college teachers as judged by teachers themselves, current and former students colleagues, administrators and external (neutral) observers. *Research in Higher Education*, 30, 137-194.
- Franklin, J. L., & Theall, M. (1989, April). *Who reads ratings: knowledge attitude and practice of users of student ratings of instruction*. Paper presented at the 70th Annual meeting of the American Educational Research Association, San Francisco, CA.
- Gervais, R., Nadeau, G., & Barnabé, C. (1996). Validation d'un instrument pour mesurer les attitudes des enseignants envers leur évaluation. *Mesure et évaluation en éducation*, 19(1), 1-24.
- Gray, D., & Brandenburg, D. (1985). Following student ratings over time with a catalog-based system. *Research in Higher Education*, 22(2), 155-168.
- Glascock, J., & Ruggiero, T. (2006). The relationship of ethnicity and sex to professor credibility at a culturally diverse university. *Communication Education*, 55, 197- 207.
- Greenwald, A. G., & Gillmore, G. M. (1997). No pain, no gain? the importance of measuring course workload in student ratings of instructions. *Journal of Educational Psychology*, 89(4), 743-751.
- Hampton, S., & Reiser, R. (2004). Effects of a theory-based feedback and consultation process on instruction and learning in college classroom. *Research in Higher Education*, 45(5), 497-526.
- Heine, P., & Maddox, N. (2009). Student perceptions of the faculty course evaluation process: An exploratory study of gender and class differences. *Research in Higher Education Journal*, 3, 1-10.
- Howard, G. S., & Maxwell, S. E. (1982). Do grades contaminate students' evaluation of college teaching: A multi-trait multimethod analysis. *Journal of Educational Psychology*, 74, 264-279.
- Jenkins, J. G. (1946). Validity for what? *Journal of Consulting Psychology*, 10, 93-98.
- Kogan, L. R., Schoenfeld-Tacher, R., & Hellyer, P. (2010). Student evaluations of teaching: Perceptions of faculty based on gender, position and rank. *Teaching in Higher Education*, 15(6), 623-636.

- Kulik, J. A., & McKeachie, W. J. (1975). The evaluation of teachers in higher education. In F. N. Kerlinger (Ed.), *Review of research in education*. Itasca, IL: Peacock ed.
- Le Boterf, G. (2005). *Ingénierie et évaluation des compétences*. Paris: Éditions d'organisation.
- Lucas, A. F. (1990). The department chair as change agent. In P. Seldin (Ed.), *How administrators can improve teaching, moving from talk to action in higher education* (pp. 63-88). San Francisco, CA: Jossey-Bass.
- Marsh, H. W. (1980). The influence of student, course, and instructor characteristics in evaluations of university teaching. *American Educational Research Journal*, 17(2), 219-237.
- 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.
- Marsh, H. W. (1984). Students evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76(5), 707-754.
- Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11(3), 255-388.
- Marsh, H. W. (2001). Distinguishing between good (useful) and bad workload on students' evaluations of teaching. *American Educational Research Journal*, 38(1), 183-212.
- Marsh, H. W., & Cooper, T. (1981). Prior subject interest, students' evaluations, and instructional effectiveness. *Multivariate Behavioral Research*, 16, 82-104.
- Marsh, H. W., Fleiner, H., & Thomas, C. S. (1975). Validity and usefulness of student evaluations of instructional quality. *Journal of Educational Psychology*, 67, 833-839.
- Marsh, H. W., & Hattie, J. (2002). The relationship between research productivity and teaching effectiveness: Complementary, antagonistic, or independent constructs?. *The Journal of Higher Education*, 73(5), 604-641.
- Marsh, H. W., & Ware, J. E. (1982). Effects of expressiveness, content coverage, and incentive on multidimensional student rating scale: New interpretations of the Doctor Fox effect. *Journal of Educational Psychology*, 74, 126-134.
- Morgan, D. A., Sneed, J., & Swinney, L. (2003). Are student evaluation a valid measure of teaching effectiveness: perceptions of accounting faculty members and administrators. *Management Research News*, 26(7), 17-32.
- Murray, H. G. (1983). Low-inference classroom teaching behaviors and student ratings of college teaching effectiveness. *Journal of Educational Psychology*, 75, 138-149.
- Naftulin, D. H., Ware, J. E., & Donnelly, F. A. (1973). The Doctor Fox lecture: A paradigm of educational seduction. *Journal of Medical Education*, 48, 630-635.
- Nasser, F., & Fresko, B. (2002). Faculty views on student evaluation of college teaching. *Assessment and Evaluation in Higher Education*, 27, 187-198.
- Ory, J. C., Braskamp, L. S., & Pieper, D. M. (1980). Congruency of student evaluative information collected by three methods. *Journal of Educational Psychology*, 72, 321-325.
- Overall, J. U., & Marsh, H. W. (1980). Students' evaluations of instruction: A longitudinal study of their stability. *Journal of Educational Psychology*, 72, 181-185.

- Paulsen, M. B., & Feldman, K. A. (1995). Toward a reconceptualization of scholarship: A human action system with functional imperatives. *The Journal of Higher Education*, 66(6), 615-640.
- Penny, A., & Coe, R. (2004). Effectiveness of consultation on student ratings feedback: a meta-analysis. *Review of Educational Research*, 74(2), 215-253.
- Perrenoud, P. (1997). *Construire des compétences dès l'école*. Paris : ESF.
- Piéron, H. (1963). *Examens et docimologie*. Paris : Presse Universitaire de France.
- Prégent, R., Bernard, H., & Kozanitis, A. (2009). *Enseigner à l'université dans une approche-programme*. Montréal : Presses internationales Polytechnique.
- Renaud, R. D., & Murray, H. G. (1996). Aging, personality, and teaching effectiveness in academic psychologists. *Research in Higher Education*, 37, 323-340.
- Rey, B., Carette, V., Defrance, A., & Kahn, S. (2006). *Les compétences à l'école : apprentissage et évaluation*. Bruxelles : De Boeck.
- Roegiers, X. (2001). *Une pédagogie de l'intégration : compétences et intégration des acquis dans l'enseignement*. Bruxelles : De Boeck.
- Romainville, M., & Coggi, C. (2009). *L'évaluation de l'enseignement par les étudiants. Approches critiques et pratiques innovantes*. Bruxelles : DeBoeck.
- Sall, C. T. (2009). Connaître les perceptions des enseignants à propos de l'évaluation des enseignements, une étape dans le processus de mise en œuvre d'une démarche qualité. In M. Romainville, & C. Coggi (Eds.), *L'évaluation de l'enseignement par les étudiants. Approches critiques et pratiques innovantes*. (pp. 213-233). Bruxelles : DeBoeck.
- Schuckman, H. (1990). Students' perception of faculty and graduate students as classroom teachers. *Teaching of Psychology*, 17(3), 162-165.
- Seldin, P. (1993). The use and abuse of student ratings of instruction. *The Chronicle of Higher Education*, 21, 40.
- Sireci, G. (2009). Packing and unpacking sources of validity evidence. In W. Lissitz (Ed.), *The concept of validity: revisions, new directions, and applications* (pp. 19-37). Charlotte, NC: Information Age.
- Sojka, J., Gupta, A. K., & Deeter-Schmelz, D. R. (2002). Student and faculty perceptions of student evaluations of teaching: A study of similarities and differences. *College Teaching*, 50(2), 44-49.
- Stufflebeam, D. (1980). *L'évaluation et la prise de décision en éducation*. Victoriaville, Canada : N.H.P.
- Tardif, J. (2006). *L'évaluation des compétences. Documenter le parcours de développement*. Montréal : Chenelière Éducation.
- Thivierge, A., & Bernard, H. (1996). Les croyances des étudiants à l'égard de l'évaluation de l'enseignement. *Mesure et évaluation en éducation*, 18(3), 59-84.
- Wright, W. A., & O'Neil, M. C. (1995). Teaching improvement practices: International perspectives. In W. A. Wright (éd.), *Teaching improvement practices. Successful strategies for higher education* (pp. 1-57). Bolton: Anker Publishing.
- Younès, N. (2006). *L'effet évaluation de l'enseignement supérieur par les étudiants* (Thèse de doctorat non publiée). Université de Grenoble, Grenoble.
- Younès, N. (2007). À quelles conditions l'évaluation formative de l'enseignement par les étudiants est-elle possible en France? *Revue française de pédagogie*, 161, 25-40.

Date de réception : 31 mai 2012

Date de réception de la version finale : 8 octobre 2012

Date d'acceptation : 11 octobre 2012