



ORGANISATION INTERNATIONALE DU TRAVAIL
ORGANISATION DES NATIONS UNIES
POUR L'ÉDUCATION, LA SCIENCE ET LA CULTURE



CEART/13/2018/4

Comité conjoint OIT/UNESCO d'experts sur l'application des Recommandations concernant le personnel enseignant (CEART)

**L'évaluation de l'enseignement dans l'enseignement supérieur:
problèmes, tendances et recommandations**

Amanda Jarrell
Université McGill
Juin 2018

Rapport présenté à la 13^e session du Comité conjoint OIT-UNESCO d'experts
sur l'application des Recommandations concernant le personnel enseignant (CEART)

Genève, 2018



Cet ouvrage est publié en libre accès sous la licence Creative Commons Attribution 3.0 IGO License (<http://creativecommons.org/licenses/by/3.0/igo>). Les utilisateurs sont autorisés à réutiliser, partager ou adapter la publication originale ou à s'en servir pour créer un nouveau produit, même à des fins commerciales, conformément aux termes de ladite licence. L'OIT et l'UNESCO doivent être reconnues comme propriétaires de l'ouvrage original. L'utilisation du logo de l'OIT, du BIT, et/ou de l'UNESCO n'est pas autorisée dans le cadre du travail de l'utilisateur.

Traductions – dans le cas d'une traduction, le déni de responsabilité suivant doit apparaître avec la mention de la source de l'ouvrage original: *Cette traduction n'a pas été réalisée par le BIT ou l'UNESCO et ne doit pas être considérée comme une traduction officielle de ce dernier. Le BIT n'est pas responsable du contenu ou de l'exactitude de la traduction.*

Adaptations – dans le cas d'une adaptation, le déni de responsabilité suivant doit apparaître avec la mention de la source de l'ouvrage original: *Cet ouvrage est une adaptation d'une publication originale du BIT et de l'UNESCO. Les idées et opinions exprimées dans cette adaptation n'engagent que la responsabilité de son auteur ou de ses auteurs et en aucun cas celle du BIT ou de l'UNESCO.*

Toute question concernant les droits et licences devra être envoyée à Publications du BIT (Droits et licences), CH-1211 Genève 22, Suisse, ou par courriel à rights@ilo.org.

L'évaluation de l'enseignement dans l'enseignement supérieur: problèmes, tendances et recommandations, par
Amanda Jarrell

Ce rapport a servi de document de référence dans le cadre de la 13e session du Comité conjoint OIT/UNESCO d'experts sur l'application des Recommandations concernant le personnel enseignant (CEART), qui s'est tenue à Genève du 1er au 5 octobre 2018.

Il est également disponible en anglais: *Assessing teaching in higher education: Current issues, trends and recommendations*

Les désignations utilisées dans les publications du BIT et de l'UNESCO, qui sont conformes à la pratique des Nations Unies, et la présentation des données qui y figurent n'impliquent de la part du BIT ou de l'UNESCO aucune prise de position quant au statut juridique de tel ou tel pays, zone ou territoire, ou de ses autorités, ni quant au tracé de ses frontières.

Les articles, études et autres textes signés n'engagent que leurs auteurs, et leur publication ne signifie pas que le BIT ou l'UNESCO souscrivent aux opinions qui y sont exprimées.

La mention ou la non-mention de telle ou telle entreprise ou de tel ou tel produit ou procédé commercial n'implique de la part du BIT ou de l'UNESCO aucune appréciation favorable ou défavorable.

Cette publication a été réalisée par le Service de production, impression et distribution des documents et publications (PRODOC) du BIT.
Création graphique, conception typographique, mise en pages, préparation de manuscrits, lecture et correction d'épreuves, impression, édition électronique et distribution.
PRODOC veille à utiliser du papier provenant de forêts gérées d'une façon qui est respectueuse de l'environnement et socialement responsable.
Code: DTP-CAD-ICA

Table des matières

	<i>Page</i>
Résumé	1
Tendances actuelles de l'enseignement supérieur	1
Introduction à l'évaluation de l'enseignement dans l'enseignement supérieur	2
Enseigner efficacement dans l'enseignement supérieur: définition	3
Principes d'évaluation	5
Validité, fiabilité, équité.....	5
Evaluer l'enseignement: méthodes et instruments de mesure	7
L'évaluation de l'enseignement par les étudiants	7
L'observation par les pairs	9
Les dossiers professionnels des enseignants	11
Les acquis des étudiants	11
Tendances et questions émergentes.....	13
La précarisation du corps enseignant	13
L'apprentissage électronique et l'enseignement en ligne.....	14
Résumé des recommandations	15
Domaines de recherche future	16
Conclusions	17
Bibliographie	19

Résumé

Evaluer l'enseignement, afin d'identifier les possibilités de perfectionnement des enseignants et de mesurer leur efficacité, est indispensable pour assurer un apprentissage et un enseignement de qualité dans l'enseignement supérieur. Pour ce faire, les systèmes d'évaluation doivent assurer l'accès à des données fiables, pouvant donner lieu à des interprétations valides et équitables de l'enseignement efficace. Les évaluations de l'enseignement doivent aussi répondre aux demandes et aux défis nouveaux créés par la massification de l'éducation et l'application croissante des mécanismes d'assurance-qualité dans l'enseignement supérieur. La présente étude a pour but d'examiner ces défis et problèmes et de proposer des recommandations en vue de promouvoir des évaluations de qualité. On s'y interroge en particulier sur les défis posés par l'absence de normes universelles en matière de conceptualisation et de mesure de l'enseignement efficace et sur les avantages et les inconvénients des méthodes d'évaluation en vigueur. Sont également examinés des défis émergents pour l'évaluation de l'enseignement que sont la précarisation des personnels et le développement de l'enseignement en ligne. Plusieurs recommandations stratégiques sont formulées, portant, notamment, sur l'élaboration de normes d'évaluation cohérentes, l'accès à la formation en matière de mesure et d'évaluation et l'amélioration de la transparence des pratiques d'évaluation.

Tendances actuelles de l'enseignement supérieur

Ces dernières décennies, l'enseignement supérieur n'a cessé de se développer au niveau mondial. Selon les estimations de l'Institut de statistique de l'UNESCO (ISU), on comptait environ 32,5 millions d'étudiants inscrits en 1970. Ils étaient près de 178 millions en 2010 et devraient approcher 263 millions en 2025 (Altbach *et al.*, 2009). Cette croissance s'est réalisée par étapes dans les différentes régions: le Canada et les Etats-Unis ont été les premiers à massifier l'enseignement supérieur dans les années soixante, suivis, dès les années quatre-vingt, par l'Europe de l'Ouest et le Japon (Altbach *et al.*, 2009). Ces quarante dernières années, l'expansion de l'enseignement supérieur a été plus marquée dans les régions émergentes, y compris en Afrique subsaharienne, dans les Etats arabes et en Amérique latine et dans les Caraïbes, avec des taux de croissance moyens oscillant entre 6 et 8,4 pour cent (Altbach *et al.*, 2009). Une croissance plus récente et d'ampleur phénoménale s'observe dans l'enseignement supérieur chinois: entre 2011 et 2016, le taux brut de scolarisation au niveau tertiaire est passé de 25,29 pour cent à 48,44 pour cent (ISU).

Cette croissance a provoqué l'augmentation mondiale du nombre d'établissements, des dépenses publiques et privées, de l'accès des jeunes aux universités et des taux d'achèvement des études (OCDE, 2017). En Inde, par exemple, le système d'enseignement supérieur comptait 27 universités en 1949 et 361 en 2006 (Altbach *et al.*, 2009). En Chine, le nombre d'écoles supérieures et d'universités a plus que doublé entre 2000 et 2010, passant de 1 041 à 2 358 établissements, grâce, essentiellement, à la croissance de l'enseignement privé (Stanfield et Shimmi, 2014). Les gouvernements ont aussi adopté des politiques d'élargissement de l'accès. Le Ghana, le Kenya, l'Ouganda et la République-Unie de Tanzanie, par exemple, ont abaissé les critères d'admission des femmes dans l'enseignement supérieur afin d'accroître leur accès et leur participation (Altbach *et al.*, 2009). D'autres pays, comme les Etats-Unis ou le Royaume-Uni, pourraient continuer de favoriser l'expansion de l'enseignement supérieur en élargissant l'accès ou en relevant le niveau d'études des adultes (Vincent-Lancrin, 2008). Dans certains pays, la massification de l'enseignement supérieur s'est également accompagnée d'une augmentation des dépenses publiques. Ainsi, les dépenses fédérales australiennes dans l'éducation tertiaire sont passées de 1,2 pour cent du produit intérieur brut (PIB) en 2010 à 1,4 pour cent du PIB en 2014 (Banque mondiale: Statistiques de l'éducation de la Banque mondiale). Cela est dû au fait

que, dans de nombreux pays, les établissements d'enseignement supérieur du secteur public reçoivent des subventions en fonction du nombre d'étudiants inscrits ou du nombre de diplômés qu'ils produisent (Santiago *et al.*, 2008).

Cette massification de l'enseignement supérieur a amené à s'interroger sur la qualité de l'enseignement dispensé (OCDE, 2017) et a suscité la création et le développement d'organismes d'assurance-qualité et d'accréditation, au point que «l'assurance-qualité dans l'enseignement supérieur est désormais une priorité pour de nombreux pays» (Altbach *et al.*, 2009, p. viii). Une définition générale de l'assurance-qualité en fait implicitement le «processus qui consiste à convaincre la partie prenante que l'offre (intrants, processus et extrants) respecte des attentes ou satisfait à un ensemble d'exigences minimales» (Harvey, 2018). Aux Etats-Unis, par exemple, les accréditations et les normes sont utilisées pour assurer une éducation de qualité dans l'enseignement supérieur (National Advisory Committee on Institutional Quality and Integrity, 2011). De plus en plus, on demande aux établissements de l'enseignement supérieur de fournir les preuves de leur fiabilité en tant que prestataires d'éducation, pour garantir aux employeurs, aux décideurs et aux autres parties prenantes que tous les étudiants bénéficient d'une éducation de qualité. Les établissements sont notamment tenus de soutenir différents types d'apprenants, aux besoins, capacités et horizons divers. Les mécanismes de contrôle de la qualité sont variables d'un pays à l'autre. D'après son site Internet, le Réseau international des organismes d'assurance de la qualité dans l'enseignement supérieur (RIOPQES) compte 187 membres titulaires, représentant 93 pays. Dans certains pays, le contrôle est centralisé. En Australie, par exemple, c'est l'Agence pour la qualité et la normalisation de l'éducation tertiaire (TEQSA), un organisme indépendant rattaché au gouvernement fédéral, qui veille à la qualité de l'enseignement et de l'apprentissage dans l'enseignement supérieur. De même, en Chine, le ministère de l'Éducation a lancé le projet Qualité et réforme de l'enseignement et de l'apprentissage dans le premier cycle de l'enseignement supérieur afin d'améliorer la qualité de l'enseignement et de l'apprentissage à ce niveau. Aux Etats-Unis, des organismes régionaux et nationaux sont chargés de contrôler la qualité.

Dans l'enseignement supérieur, les mécanismes d'assurance-qualité exigent des établissements qu'ils conçoivent un enseignement efficace et en apportent les preuves. Autrement dit, les établissements doivent avoir les moyens de reconnaître et de mesurer la qualité des pratiques didactiques mises en œuvre entre leurs murs et les domaines à améliorer, en évaluant les intrants, les processus et les extrants de l'éducation. Cela pose de nouveaux défis aux établissements, dans la mesure où, traditionnellement, on s'est surtout appuyé sur leur productivité en matière de recherche pour juger de l'efficacité des personnels universitaires, et où «le secteur de l'enseignement supérieur n'est généralement pas bien préparé à être davantage tenu pour responsable de ses résultats éducatifs» (Lodge et Bonsanquet, 2014, p. 5). Or tant la massification de l'enseignement que l'obligation de rendre des comptes font de l'évaluation de l'enseignement une importante nécessité et exigent impérativement la création de mécanismes d'évaluation transparents et de qualité. Dans les sections qui suivent, nous passerons en revue les problèmes que cela suscite et fournirons des recommandations sur la réalisation de cet objectif.

Introduction à l'évaluation de l'enseignement dans l'enseignement supérieur

La mesure et l'évaluation de la performance enseignante sont cruciales pour dispenser une éducation de qualité et mettre celle-ci en évidence. «Il est essentiel de connaître les points forts des enseignants et les aspects de leur pratique qui doivent être perfectionnés. De ce point de vue, la mise en place d'une évaluation des enseignants est une étape cruciale dans les efforts pour améliorer l'efficacité de l'enseignement et relever la qualité de l'éducation.» (OCDE, 2009) Les évaluations de l'enseignement supérieur peuvent s'effectuer à différents niveaux et cibler les individus, les programmes ou l'établissement.

Évaluer chacun de ces niveaux est essentiel pour contrôler et préserver la qualité globale de l'éducation. Les paramètres d'ordre général utilisés pour évaluer la qualité de l'enseignement au niveau des établissements portent le plus souvent sur les intrants et les processus – tels que les taux d'encadrement ou la durée d'achèvement de la formation – et ne sont pas étroitement corrélés à l'apprentissage des étudiants (Salmi et Saroyan, 2007; Strang *et al.*, 2016), alors que, lorsqu'elle est mesurée au niveau de l'enseignant, la qualité de l'enseignement est souvent associée à l'apprentissage. Les évaluations au niveau individuel ont aussi pour autre intérêt d'être utilisées dans les décisions concernant le personnel, comme le recrutement ou la titularisation, et de permettre d'identifier les possibilités de développement professionnel. L'examen des évaluations de l'enseignement individuel, ou même d'un ensemble de résultats d'enseignement individuels, peut donner une meilleure idée de la qualité de l'enseignement et de l'apprentissage. Les principes examinés plus bas s'appliquent aux évaluations effectuées à tous les niveaux. Toutefois, compte tenu de l'importance des évaluations de l'enseignement effectuées au niveau individuel, c'est surtout à ce dernier aspect que s'attachera la présente étude.

L'évaluation de l'enseignement dans l'enseignement supérieur est généralement utilisée à des fins d'amélioration et de reddition des comptes et comme critère pour permettre aux étudiants de sélectionner une formation. Les évaluations de l'enseignement (y compris celles opérées par les étudiants, également appelées «évaluations de cours») ont d'abord été introduites à des fins de perfectionnement. Ce type d'évaluation est généralement de nature formative et utilisé pour fournir aux enseignants des commentaires utiles leur permettant d'améliorer la qualité de leur pratique et d'identifier leurs besoins en développement professionnel. L'évaluation à des fins de reddition des comptes est généralement sommative et sert à tenir les enseignants pour responsables de la qualité de leur enseignement. Ses résultats sont utilisés pour prendre des décisions concernant les personnels, comme la rétention, le renouvellement, la promotion, ou même les augmentations de salaire au mérite. Si des méthodes similaires peuvent être utilisées à des fins d'amélioration de l'enseignement ou de reddition des comptes, appliquer la même méthode pour atteindre ces deux objectifs peut engendrer des tensions et poser des difficultés particulières.

Dans les sections suivantes, nous aborderons les problèmes posés par l'évaluation de l'enseignement dans l'enseignement supérieur. Nous examinerons d'abord différentes conceptualisations de l'enseignement efficace et leur influence sur la conception des outils d'évaluation. Nous donnerons ensuite un aperçu des principes d'évaluation et de la manière dont ils peuvent guider les pratiques. Puis, nous analyserons les avantages et les inconvénients des méthodes et instruments d'évaluation qui ont cours actuellement. Enfin, nous évoquerons quelques-unes des tendances et des questions émergentes concernant l'évaluation de l'enseignement dans l'enseignement supérieur. Un résumé de nos recommandations et suggestions de recherches futures conclura cette étude.

Enseigner efficacement dans l'enseignement supérieur: définition

Dans l'enseignement supérieur, l'enseignement est un processus extrêmement complexe comprenant, en gros, «la conception des programmes, le choix des contenus et des méthodes, les diverses formes d'interaction entre les enseignants et les étudiants et l'évaluation de ces derniers» (Ramsden, 2003, p. 85). L'approche qu'ont les enseignants de l'enseignement est influencée par leurs conceptions de l'enseignement et de l'apprentissage et leurs perceptions de l'environnement d'enseignement (Prosser et Trigwell, 1997; Trigwell et Prosser, 1996). Une des principales questions auxquelles les évaluations de l'enseignement dans l'enseignement supérieur tentent de répondre est: «L'enseignement dispensé par les enseignants est-il réellement (et suffisamment) efficace?» Pour y répondre, il faut disposer de normes de référence fiables ou d'une définition claire des compétences d'enseignement attendues dans le cadre de l'enseignement supérieur. Cela mène à des

questions plus générales concernant la qualité et l'efficacité de l'enseignement dans l'enseignement supérieur.

Il n'existe pas de normes universellement acceptées d'enseignement efficace et cette question reste controversée dans la littérature (Skelton, 2005; Kember *et al.*, 2006). On peut définir largement l'enseignement efficace en fonction des activités de l'enseignant survenant avant (préparation) et pendant (exécution) l'acte d'enseigner et qui produisent des changements pertinents chez les élèves (Abrami *et al.*, 2007). Les attentes précises et les exigences minimales en matière d'enseignement efficace ne sont pas clairement définies. Il est généralement admis que les compétences et les pratiques des enseignants efficaces dépendent du contexte dans lequel se déroulent l'enseignement et l'apprentissage. Par conséquent, il n'existe pas un seul, mais plusieurs processus d'enseignement efficace. En outre, l'enseignement efficace ne donne pas lieu à un seul type de résultats et la valeur attribuée à ces derniers dépend du contexte. Des ajustements sont en effet nécessaires pour répondre aux besoins locaux, tels qu'ils auront été décrits par les différentes parties prenantes: «étudiants, employeurs, personnel enseignant et non enseignant, gouvernement et organismes de financement, créanciers, commissaires aux comptes, et la collectivité dans son ensemble» (Hénard, 2010, p. 87). L'enseignement efficace est donc contextuel et pluridimensionnel et se compose d'actes multiples et éventuellement distincts (Abrami *et al.*, 2007).

Les données concernant la manière dont les établissements dans les différentes régions définissent l'enseignement efficace sont difficiles à obtenir et on manque actuellement d'une analyse comparative globale des normes et exigences institutionnelles en la matière. Les publications sur ce sujet offrent un certain aperçu des points de vue sur ce qui constitue le «bon» enseignement dans l'enseignement supérieur. Une des premières orientations dans ce domaine est apportée par les sept principes de Chickering et Gamson (1987), qu'on trouve cités plus de 7 000 fois. Selon eux, l'enseignement efficace dans le premier cycle de l'enseignement supérieur devrait obéir aux principes suivants: encourager les contacts entre le corps enseignant et les étudiants; favoriser la réciprocité et la coopération entre les étudiants; appliquer des techniques d'apprentissage actif; fournir rapidement des commentaires en retour; mettre l'accent sur le temps consacré aux tâches; communiquer des attentes élevées; respecter la diversité des talents et des façons d'apprendre. Bien que cette publication date de plus de trente ans, la plupart de ses préconisations continuent d'être mises en avant (par exemple, par Tobin *et al.*, 2015). Plusieurs mettent l'accent sur les actions ou les comportements de l'enseignant pouvant encourager l'apprentissage des étudiants. Un modèle plus récent d'enseignement, proposé par Saroyan et Amundsen (2001), prend en compte, outre l'acte d'enseigner, les croyances et les connaissances, considérant ces trois entités comme des éléments clés pour dispenser un enseignement efficace. Ce modèle a été conçu au départ à des fins d'amélioration des personnels et, à ce titre, il souligne l'importance d'un alignement entre ses principales composantes. Les croyances concernant l'enseignement et l'apprentissage peuvent comprendre des convictions individuelles concernant les rôles et les responsabilités des enseignants et des étudiants au sein d'un contexte d'enseignement particulier et les éléments constitutifs d'un enseignement et un apprentissage efficaces. Quant à l'élément de connaissances, il désigne celles qui se rapportent à la matière enseignée, à la pédagogie et aux besoins des étudiants. Enfin, les actions, selon ce modèle, englobent les processus et les résultats de l'enseignement. Tous ces éléments interagissent les uns avec les autres au sein du contexte d'instruction particulier (par exemple, niveau, type de cours, discipline, etc.), et la définition de l'enseignement efficace proposée dans ce modèle est donc plus dynamique que statique.

Les mises en œuvre effectives de l'enseignement efficace fournissent des informations sur les données qui seront collectées et acceptées en vue d'évaluer l'enseignement. Car, «une fois définies les notions de qualité et d'enseignement, l'établissement est en meilleure position pour déterminer les instruments les mieux adaptés pour évaluer cette qualité» (Hénard, 2010, p. 87). Les indicateurs d'enseignement efficace dans l'enseignement

supérieur et les moyens de le mesurer dans ce contexte restent obscurs. L'absence de normes admises d'enseignement efficace empêche d'élaborer des pratiques cohérentes d'évaluation des enseignants dans l'enseignement supérieur. Comme recommandé par Strang *et al.* (2016), «il faut parvenir à une compréhension commune de la notion d'«enseignement de qualité» dans l'enseignement supérieur, des buts et priorités des [établissements d'enseignement supérieur] eux-mêmes, et de la façon dont ces éléments doivent être harmonisés» (p. 7). On peut en trouver un parallèle dans le secteur de l'éducation postsecondaire, où de nombreuses régions ont défini des normes professionnelles ou précisé les compétences requises pour enseigner, qui sont utilisées pour soutenir la formation professionnelle, l'autoévaluation, la réflexion et l'évaluation (voir, par exemple, les *Normes d'exercice de la profession enseignante* de la province canadienne de l'Ontario). Au niveau de l'enseignement supérieur, un consensus entre parties prenantes et décideurs sur la définition de l'enseignement de qualité et les types de données qu'il convient de recueillir et d'accepter comme indicateurs d'enseignement efficace aurait son utilité. Dans un contexte où l'internationalisation va croissant, il y a un réel besoin de poser pour principe que les acquis des diplômés sont comparables et que la qualité de l'enseignement est une condition essentielle de l'apprentissage des étudiants. Or des questions demeurent concernant la flexibilité requise pour rendre compte des objectifs et des besoins des contextes locaux. Les établissements devraient-ils, par exemple, définir avant tout l'enseignement efficace en fonction de leur contexte particulier en termes d'intrants, de processus et d'extrants? Ces définitions devraient-elles refléter le caractère pluridimensionnel de l'enseignement, et les établissements, dans un effort de transparence, ne devraient-ils pas envisager de les mettre à disposition, notamment des parties prenantes?

Principes d'évaluation

Que sa cible soit les intrants, les processus ou les extrants, l'évaluation de l'enseignement doit produire des résultats fiables. C'est un principe fondamental car, si la qualité des mesures d'évaluation employées est mauvaise, les conclusions tirées auront peu de valeur. Cela aura ensuite des conséquences non négligeables sur certaines décisions critiques, par exemple en ce qui concerne les personnels, et sur les conclusions concernant la qualité de l'éducation. Les principes régissant l'élaboration et l'appréciation des évaluations éclairent sur cette question essentielle. L'American Psychological Association a défini une série de principes, sous le titre *Standards for educational and psychological testing* [Normes d'évaluation pédagogiques et psychologiques] (American Psychological Association, 2014). Ces normes fournissent notamment des consignes pour l'utilisation des évaluations dans les décisions concernant la sélection, l'affectation et la promotion des personnels. Nous nous en servons dans la présente étude pour apprécier les méthodes aujourd'hui utilisées pour évaluer l'enseignement dans l'enseignement supérieur et pour formuler des recommandations. Dans les prochaines sections, nous passerons brièvement en revue les principaux principes examinés dans ces *Standards*: la validité, la fiabilité et l'équité. Généralement parlant, ces principes permettent de porter un jugement objectif sur les évaluations et d'obtenir des résultats exacts et fiables.

Validité, fiabilité, équité

Les trois principes fondamentaux de l'évaluation sont la validité, la fiabilité et l'équité. La validité est la mesure dans laquelle les faits et la théorie permettent d'interpréter les résultats de l'évaluation pour l'usage qui en est proposé (American Psychological Association, 2014). En d'autres termes, la validité se rapporte au bien-fondé, à la véracité ou à la légitimité des conclusions tirées à partir des résultats de l'évaluation. En ce qui concerne l'évaluation de l'enseignement dans l'enseignement supérieur, cela équivaut aux allégations concernant l'efficacité ou la qualité de l'enseignement. Le principe de fiabilité renvoie à la relative précision des résultats. Il prend de l'importance à mesure que celle des

interprétations augmente. Ainsi, si la mesure d'évaluation est utilisée pour prendre des décisions sur lesquelles il est difficile de revenir (comme les décisions concernant le personnel, telles que le recrutement ou la titularisation), un haut degré de fiabilité est alors souhaitable. Plus la mesure est fiable, moins on s'expose au risque d'une erreur de mesure. La fiabilité détermine donc notre degré de confiance dans les résultats d'évaluation et, par extension, la véracité des interprétations fondées sur ces résultats. La fiabilité peut être influencée par de nombreux facteurs, comme les changements dans la procédure d'évaluation, les modifications apportées à l'évaluation elle-même, les différences entre les évaluateurs et les différences entre les populations. Par conséquent, il est important de vérifier qu'une évaluation de l'enseignement conçue pour un contexte en particulier restera fiable dans un autre contexte. Enfin, le principe d'équité fait référence à une évaluation qui «n'avantage ni ne désavantage certains individus du fait de caractéristiques sans rapport avec le construit visé» (American Psychological Association, 2014, p. 50), comme la race, l'origine ethnique, le genre, l'âge, la classe sociale, ou la langue ou la culture d'origine. Lorsque certains aspects de l'évaluation ou de la procédure d'évaluation «abaissent ou relèvent systématiquement les scores de groupes identifiables participant aux évaluations, [cela peut conduire à] des interprétations de scores non conformes aux usages prévus» (American Psychological Association, 2014, p. 54). L'équité doit être recherchée à toutes les étapes de la conception et de l'administration des évaluations, ainsi que de l'interprétation de leurs résultats.

Des données attestant la fiabilité, la validité et l'équité de l'évaluation doivent être systématiquement recueillies et documentées pour garantir que les résultats sont exacts et les interprétations dignes de foi. Plusieurs types de données peuvent être collectées comme preuves de validité, telles que les preuves basées sur le contenu de l'évaluation et les relations avec d'autres variables. Les preuves de validité basées sur le contenu doivent démontrer qu'il y a alignement entre le contenu de l'évaluation et le construit mesuré. Les preuves basées sur les relations avec d'autres variables doivent démontrer que les résultats de l'évaluation sont en lien avec d'autres variables dans le sens et avec l'intensité attendus. On peut s'attendre à ce que les résultats de l'évaluation soient en étroite corrélation, ou au contraire sans lien, avec d'autres variables. De même, pour attester la fiabilité de l'évaluation, il faut des données probantes, comme la preuve que l'instrument d'évaluation peut fournir des résultats stables dans le temps. Les preuves d'équité sont plus difficiles à établir, parce que «les différences de résultats entre les groupes ne démontrent pas par eux-mêmes qu'un test est biaisé ou inéquitable» (American Psychological Association, 2014, p. 54). Toute différence de résultats entre deux groupes exige un examen plus approfondi pour comprendre la cause de cet écart. A titre d'exemple, les résultats d'une évaluation de l'enseignement pourraient indiquer que les enseignants en fin de carrière sont moins performants que les enseignants débutants. Il se peut que cette évaluation, involontairement, désavantage les professeurs expérimentés et avantage les novices. Mais il se peut aussi que les différences soient dues à de réelles différences d'efficacité ou de qualité des enseignements. C'est pourquoi les différences basées sur des groupes identifiables devraient être examinées avec soin, et, lorsqu'un parti pris réel est détecté, des procédures doivent être prévues pour rétablir l'équité. L'élaboration, l'administration et l'interprétation des évaluations de l'enseignement devraient obéir à des normes de mesure et d'évaluation, telles que celles définies par l'American Psychological Association ou tout autre organisme équivalent. Ces normes de référence fournissent une approche systématique des évaluations de l'enseignement capables de soutenir des pratiques de qualité. Des preuves de fiabilité, de validité et d'équité devraient être recueillies et documentées pour que les résultats permettent de statuer avec précision et infailibilité sur la qualité de l'enseignement dispensé. Les établissements devraient aussi fournir une documentation indiquant la bonne manière d'utiliser l'évaluation et de l'interpréter. Celle-ci contiendra des informations concernant la nature de l'évaluation, l'usage auquel elle est destinée, la façon dont elle a été élaborée, les méthodes de notation et d'interprétation et la collecte de preuves de validité, de fiabilité et d'équité (American Psychological Association, 2014). Tout cela est important à la fois pour les évaluations conçues localement pour diffusion au sein des établissements et pour celles

qui sont élaborées en externe et administrées par une université en particulier. Pour promouvoir des pratiques d'évaluation défendables, le personnel des universités peut avoir besoin d'une formation et d'un accès aux spécialistes des évaluations.

Evaluer l'enseignement: méthodes et instruments de mesure

Comme indiqué plus haut, on ne dispose pas de normes universelles d'enseignement efficace. Aussi beaucoup d'établissements ont-ils renoncé à en chercher une définition pour se concentrer sur les moyens de le mesurer (Pitman, 2014). A titre d'exemple, dans le Cadre d'excellence de l'enseignement (TEF) du ministère des Entreprises, de l'Innovation et des Compétences du Royaume-Uni (BIS, 2015), on lit qu'«il n'existe pas de définition communément admise de l'enseignement de qualité dans l'enseignement supérieur [...]. En l'absence de toute définition approuvée ou de mesures reconnues de la qualité de l'enseignement, le gouvernement propose d'utiliser des mesures, ou des paramètres, comme indicateurs de la qualité de l'enseignement. En conséquence, la difficulté réside dans l'identification des paramètres pouvant mesurer avec le plus de fiabilité et de précision la qualité de l'enseignement, par opposition à d'autres facteurs contribuant aux résultats obtenus par les étudiants.» (pp. 5-6) Cela n'est toutefois pas nécessairement problématique en soi, selon Saroyan et Amundsen, qui estiment que «le problème ne réside pas tant dans le fait qu'il existe réellement ou non des comportements efficaces d'enseignement que dans la production d'une image partielle de l'enseignement en tentant de l'évaluer» (2001, p. 343). En d'autres termes, la validité des interprétations concernant l'enseignement est discutable lorsque l'évaluation n'est pas alignée sur l'enseignement efficace.

Un deuxième problème connexe est qu'il est rare que les preuves de fiabilité et de validité concernant les évaluations de l'enseignement soient établies empiriquement. Comme résumé par Strang *et al.* (2016) dans leur examen systématique des publications consacrées à la définition et à l'évaluation d'un enseignement supérieur de qualité au Royaume-Uni, on trouve dans cette littérature peu de preuves qui relient les indicateurs d'enseignement efficace couramment utilisés à la réussite des étudiants. Bien que le nombre de recommandations basées sur des données probantes concernant les méthodes et les instruments d'évaluation de l'enseignement soit actuellement limité, plusieurs de ces méthodes sont d'usage courant à travers le monde. Dans les sections qui suivent, nous les passerons rapidement en revue afin d'examiner leurs points forts et leurs points faibles.

L'évaluation de l'enseignement par les étudiants

L'évaluation de l'enseignement par les étudiants (EEE) est de loin la méthode la plus couramment employée pour évaluer l'enseignement, et elle est souvent utilisée pour prendre des décisions concernant le personnel – recrutement, titularisation, promotion ou augmentation du salaire (Clayson, 2013; Clayson et Haley, 2011; Jones, 2012; Stowell *et al.*, 2012). Cette méthode consiste en un questionnaire auquel les étudiants répondent par des informations d'ordre quantitatif ou qualitatif après avoir bénéficié d'un enseignement, généralement à l'issue de la formation. L'avantage de cette méthode est qu'elle est à la fois rentable et efficace pour mesurer des éléments utilisés comme indicateurs d'enseignement efficace. La perception qu'ont les étudiants de leur expérience d'apprentissage peut fournir des données précieuses, et on leur donne aussi l'importante occasion de s'exprimer. Un autre avantage est que cette méthode peut produire une représentation plus fiable et valide de la qualité de l'enseignement que toute autre méthode d'évaluation de l'enseignement (Berk, 2005, 2013; McKeachie, 1979).

Mais, bien qu'elle présente plusieurs avantages, cette méthode suscite aussi des préoccupations fondées. Des questions restent en suspens concernant la relative fiabilité des

données et la validité des interprétations qui en sont tirées. Certaines EEE sont élaborées rigoureusement selon des normes de mesure et d'évaluation, mais d'autres semblent conçues localement du point de vue du « bon sens » (Coffey et Gibbs, 2001). C'est le signe que les instruments de l'évaluation par les étudiants sont de qualité variable. Cette instabilité suscite des craintes chez les enseignants lorsque la méthode est utilisée à des fins de reddition des comptes (Simpson et Siguaw, 2000; Stein *et al.*, 2012; Surgenor, 2013). Une circonstance qui explique peut-être le caractère discutable de cet instrument est que, à l'origine, les EEE n'ont pas été conçues pour éclairer des décisions concernant le personnel (Hobson et Talbot, 2001). Le fait qu'elles soient depuis peu utilisées à des fins de reddition des comptes exige donc qu'elles soient évaluées avec la minutie nécessaire.

Les données suggèrent que les résultats des EEE dans l'enseignement supérieur peuvent conduire à des interprétations erronées de la qualité de l'enseignement. Plusieurs études menées dans divers contextes suggèrent que les perceptions qu'ont les étudiants de la qualité de l'enseignement peuvent être influencées par une série de facteurs comme le sexe de l'étudiant, le sexe de l'enseignant, le domaine d'études, le type d'établissement, l'année de la formation, l'attrait de l'enseignant et les opinions préalables le concernant (Freng et Webber, 2009; Hammonds *et al.*, 2017; Lewandowski *et al.*, 2011; MacNell *et al.*, 2015; Nargundkar et Shrikhande, 2014; Nasser-Abu, 2017; Onwuegbuzie *et al.*, 2007; Zhao et Gallant, 2012). En d'autres termes, des facteurs sans rapport avec la qualité effective de l'enseignement semblent systématiquement influencer les résultats de ces évaluations. Cela remet en cause la fiabilité et la validité de cette méthode. En outre, les conclusions portant sur les caractéristiques individuelles de l'enseignant (par exemple, le sexe) mettent également en doute l'équité de cette pratique d'évaluation, dans la mesure où elles semblent favoriser certains enseignants par rapport à d'autres. D'un autre côté, la littérature suggère aussi que les facteurs sans lien avec l'enseignement n'ont pas une influence significative sur les évaluations étudiantes (d'Apollonia et Abrami, 1997; Remedios et Lieberman, 2008; Marsh, 2007; Marsh et Roche, 1997). Ces conclusions contradictoires dans la littérature appellent à des examens locaux attentifs des EEE conformément aux normes de mesure et d'évaluation.

La littérature formule aussi des conclusions divergentes quant aux associations entre les évaluations de l'enseignement par les étudiants et l'apprentissage de ces derniers. L'enseignement efficace devrait être en relation avec l'apprentissage des étudiants. Or les preuves empiriques de ce lien sont contradictoires. Dans une méta-analyse récente de 51 études, qui examine l'association entre les EEE et l'apprentissage des étudiants, les auteurs concluent que les étudiants n'apprennent pas davantage auprès d'enseignants ayant été mieux notés par les évaluations étudiantes (Uttl *et al.*, 2017). On tient là une preuve convaincante que les évaluations de l'enseignement par les étudiants ne sont pas associées à l'apprentissage des étudiants. Mais il faut toutefois remarquer que d'autres études individuelles suggèrent qu'il existe des corrélations faibles à modérées entre les EEE et la notation ou la notation attendue (Abrami, 2001; Abrami *et al.*, 1980; Eiszler, 2002; Feldman, 1976; Greenwald et Gillmore, 1997; Stumpf et Freedman, 1979). Ces conclusions contradictoires font douter de la validité de cette méthode d'évaluation en tant que réelle mesure de l'enseignement efficace. De fait, les évaluations étudiantes de l'enseignement se limitent le plus souvent à l'acte d'enseigner, ce qui signifie que d'autres aspects de l'enseignement, comme la planification des cours ou la réflexion, sont difficilement saisis par cette méthode (Saroyan et Amundsen, 2001). Il est possible que l'image partielle de l'enseignement fournie par les évaluations étudiantes soit à l'origine de ces résultats empiriques incohérents, et parfois surprenants.

Une autre crainte est que l'utilisation des EEE puisse avoir des effets indésirables sur l'enseignement et l'apprentissage. Certains pensent que les éléments inclus dans les évaluations étudiantes pourraient aller à l'encontre des objectifs institutionnels et pédagogiques. Selon eux, certains instruments des EEE pourraient par exemple encourager une approche de l'enseignement centrée sur l'enseignant, et non sur l'étudiant (Abrami *et al.*,

2007). Par voie de conséquence, «il serait téméraire pour de jeunes enseignants espérant une titularisation de créer des environnements d'apprentissage fondés sur la participation active, qui reposent davantage sur la collaboration entre étudiants avec ou sans soutiens technologiques, malgré les preuves irréfutables que ces stratégies améliorent l'apprentissage, la rétention des connaissances et le maintien dans les filières scientifiques» (Abrami *et al.*, 2007, p. 451). Certaines données suggèrent aussi que les EEE pourraient être un frein aux innovations pédagogiques en classe, telles que l'expérimentation de nouvelles méthodes, qui comporte souvent des tâtonnements, des erreurs et des réajustements, ce qui peut donner lieu à des évaluations moins favorables (Darwin, 2017). Certains craignent aussi que les enseignants puissent être incités à adopter des programmes pédagogiques moins ambitieux pour ne pas s'exposer à une mauvaise notation des étudiants (Darwin, 2017). Il est donc important que les établissements évaluent les effets potentiellement indésirables des évaluations de l'enseignement par les étudiants, en particulier lorsqu'elles sont utilisées à des fins de reddition des comptes.

De façon générale, la littérature ne fournit pas de preuves convaincantes que les EEE permettent des appréciations fiables et exactes de l'efficacité de l'enseignement. Cette conclusion devrait inciter à la prudence lorsque cette méthode est employée pour juger l'enseignement, compte tenu, notamment, du poids que les évaluations pourraient avoir dans les décisions concernant le personnel. Pour pouvoir se fier à leurs résultats, les établissements devraient les analyser en s'interrogeant sur leur fiabilité, leur validité et leur équité, ainsi que leurs éventuels effets indésirables. Les conclusions de ces analyses leur fourniront des informations utiles pour améliorer l'instrument en vue de futures applications et favoriser des interprétations justes et équitables des résultats d'évaluation. Il peut être utile également de diffuser la stratégie et les conclusions d'analyse aux parties prenantes concernées, corps enseignant compris, pour améliorer la transparence et guider l'interprétation des résultats (Winer *et al.*, 2012). Les établissements devraient également avoir conscience que les EEE ne peuvent rendre compte que de ce que les étudiants sont capables d'observer en classe, ce qui signifie que d'autres aspects importants de l'enseignement efficace échappent aux mesures.

L'observation par les pairs

L'observation de l'enseignement par les pairs consiste à inviter un individu extérieur dans une classe pour qu'il observe la pratique. Elle fournit un angle d'approche supplémentaire pour évaluer l'efficacité de l'enseignement. On a suggéré que les pairs étaient mieux placés pour observer certains aspects de la qualité de l'enseignement, comme le plan de cours, que les étudiants ont plus de mal à reconnaître (Paulsen, 2002). Or les évaluations de l'enseignement efficace utilisant cette méthode restent limitées à ce qui est directement observable dans la salle de classe. Les observations sont généralement structurées et orientées par l'instrument d'évaluation sur une attention directe à des aspects spécifiques de la pratique didactique. L'observation par les pairs peut être utilisée à des fins d'amélioration et de reddition des comptes. Lorsque le but est l'amélioration de l'enseignement, l'observation, suggèrent les chercheurs, peut aider l'enseignant à prendre du recul, à gagner en confiance, à s'impliquer davantage et à être plus enclin à adopter des méthodes innovantes (Cosh, 1998). Mais des doutes subsistent sur l'utilisation de cette méthode d'évaluation à des fins de reddition des comptes. On dispose actuellement de peu de preuves empiriques de sa fiabilité et de sa validité, et les bonnes pratiques n'apparaissent pas clairement (Pounder *et al.*, 2016).

Des questions demeurent concernant le choix de l'observateur. En fonction du modèle, le pair observateur peut être un autre membre du corps enseignant (issu du même ou d'un autre département), un expert pédagogique ou un étudiant formé. Un inconvénient des observations menées par un enseignant du même département est le risque que s'instaure un rapport de force inégal (McMahon *et al.*, 2007) et que l'attention se détourne de la pratique

didactique pour se reporter sur le contenu enseigné (Torres *et al.*, 2017). L'observation entre enseignants de disciplines différentes peut s'avérer avantageuse dans la mesure où elle peut réduire l'intérêt porté au contenu pour concentrer l'observation sur la pratique effective de l'enseignement (Kinchin, 2005). On s'expose aussi au risque qu'«en tant que pairs observateurs, les enseignants, indépendamment de la discipline, soient mal équipés, si on ne leur fournit par une formation supplémentaire, pour évaluer et commenter l'efficacité de l'enseignement d'autrui» (Gosling, 2009, p. 7). Dans la même optique, les universitaires sont rarement des spécialistes de l'éducation et ne sont par conséquent pas nécessairement qualifiés pour évaluer l'enseignement efficace (Smith, 2012).

Le principal sujet de préoccupation en ce qui concerne l'observation de l'enseignement par les pairs est la partialité. Cela pose particulièrement problème lorsque cette méthode est utilisée à des fins de reddition des comptes. Il y a de nombreux types de parti pris ou de préjugés, bien documentés dans la recherche, qui peuvent affecter la fiabilité et la validité des résultats de cette approche d'évaluation de l'enseignement. Une des formes de parti pris potentiel se produit lorsque l'observateur «a une attente préconçue du comportement de la personne observée, [et qu']il ou elle est susceptible, au moins dans une certaine mesure, de faire correspondre ses observations réelles à ces attentes» (Muijs, 2006, p. 63). Il a été démontré que le parti pris de l'observateur influence les jugements portés sur l'efficacité de l'enseignement. Dans une étude expérimentale, par exemple, on indiquait à l'avance aux étudiants que l'enseignant invité était chaleureux ou distant. L'étude concluait que les étudiants à qui on avait dit que l'enseignant était chaleureux le notaient comme étant plus efficace que les étudiants à qui on avait dit qu'il était distant (Widmeyer et Loy, 1988). Ces attentes préconçues au sein du corps enseignant peuvent être difficiles à éviter et l'ampleur du parti pris difficile à quantifier.

Plusieurs stratégies permettent d'améliorer la fiabilité de cette méthode. On peut par exemple s'appuyer sur des observations multiples étalées dans le temps et employer plusieurs observateurs pour améliorer l'exactitude des résultats d'évaluation (Braskamp et Ory, 1994; Centra, 1993; Chism, 1999; DeZure, 1999; Muijs, 2006). Il est également recommandé de soumettre les observateurs à une formation spécialisée, bien qu'on ne sache pas bien quel type de formation est nécessaire pour que les pairs observateurs puissent rendre compte avec précision de l'efficacité de l'enseignement. Pour Muijs (2006), par exemple, «cette formation doit familiariser [l'observateur] avec l'instrument utilisé, l'accompagner soigneusement dans toutes les étapes à suivre, lui expliquer comment les données seront interprétées, et lui présenter des sources possibles de parti pris pour lui permettre de s'y soustraire. Une bonne dose de pratique est nécessaire, et celle-ci doit se poursuivre jusqu'à ce qu'on ait atteint un niveau de fiabilité suffisant.» (p. 64) On a quelques preuves que cette formation puisse être fructueuse. Par exemple, dans une étude de cas portant sur un programme d'observation de l'enseignement par les pairs à Hong-kong, Chine, des étudiants avaient été formés comme «consultants» pour fournir des commentaires sur l'enseignement à partir d'observations régulières sans jugement, du point de vue de l'étudiant (Pounder *et al.*, 2016).

Pour résumer, l'observation de l'enseignement par les pairs peut mesurer certains aspects de l'enseignement efficace difficiles à appréhender par d'autres méthodes. Des questions n'en demeurent pas moins concernant son efficacité et les meilleurs moyens de lutter contre le parti pris et d'améliorer la fiabilité. Idéalement, ces observations devraient être réalisées plusieurs fois et par plusieurs observateurs formés. Compte tenu du faible soutien empirique disponible pour apporter des preuves de fiabilité et de validité, les établissements devraient effectuer des évaluations internes de cette méthode pour en garantir la qualité. Ces questions prennent encore plus d'importance lorsque les résultats doivent être utilisés à des fins de reddition des comptes.

Les dossiers professionnels des enseignants

Les dossiers professionnels ou portefeuilles d'enseignement jouissent d'une faveur croissante comme méthode complète de présentation de données probantes d'enseignement efficace dans l'enseignement supérieur. Constitués par l'enseignant lui-même, suivant des consignes préétablies, ils contiennent différents éléments attestant la qualité de l'instruction dispensée. Ces éléments de preuve sont généralement présentés sous forme d'objets ou de documents, tels que les descriptions de cours, les notes personnelles, les autoévaluations de l'enseignement, l'énoncé des différentes fonctions occupées, les échantillons de travaux d'étudiants, etc. On peut ainsi saisir des aspects de l'enseignement efficace qui ne sont pas aisément observables dans une salle de classe. Certains suggèrent également que c'est la méthode la plus prometteuse à des fins d'amélioration, dans la mesure où elle donne aux enseignants une occasion de mener une réflexion sur leur pratique (Saroyan et Amundsen, 2001).

Par contre, lorsque l'on s'en sert à des fins de reddition des comptes, des questions se posent sur la fiabilité et la validité de cette méthode. Ainsi, les éléments de preuve apportés dans le dossier professionnel de l'enseignant peuvent être biaisés s'ils ne présentent qu'une image partielle et favorable de la qualité de l'enseignement dispensé. Il faut que la constitution des dossiers soit encadrée pour garantir que les données fournies sont représentatives de la qualité de l'enseignement évalué et de l'apprentissage des élèves (Abrami *et al.*, 2007). Cette représentativité est importante, dans la mesure où «une pratique courante, consistant à soumettre les meilleurs produits étudiants comme preuve d'efficacité pédagogique, ne permet guère de juger correctement de la manière dont les instructeurs favorisent l'apprentissage» (Abrami *et al.*, 2007, p. 393). Une autre question en suspens concerne notre capacité à évaluer objectivement ces dossiers professionnels des enseignants. La fiabilité et la validité de cette méthode peuvent être améliorées en instaurant des règles et des normes d'évaluation, en donnant des consignes pour la constitution du dossier, en formant les examinateurs et en soumettant les dossiers à l'examen de plusieurs (Seldin, 1993; Paulsen, 2002). S'agissant de la formation, par exemple, il a été recommandé que les examinateurs aient des «possibilités de se faire une idée des méthodes, des critères et des normes d'appréciation en consultant les dossiers déjà notés, que la note soit bonne ou mauvaise» (Paulsen, 2002, p. 13).

Les acquis des étudiants

Les acquis des étudiants sont un autre indicateur d'enseignement efficace. On définit ainsi ce que les étudiants doivent avoir appris et être capables de faire à l'issue des études, autrement dit les connaissances, compétences et attitudes acquises. L'efficacité de l'enseignement est un important facteur d'acquisition par l'apprentissage (Biggs, 2001). Les acquis d'apprentissage sont une plus large définition que l'exercice des compétences stipulé par les organisations professionnelles. L'importance accordée aux acquis d'apprentissage dans l'enseignement supérieur s'est accentuée ces dernières années dans le but de fournir une mesure de ce que les étudiants apprennent (Tremblay *et al.*, 2012). En général, les objectifs en la matière sont définis localement au sein des cours et des programmes, même si, plus récemment, certains ont plaidé pour des mesures internationales et cohérentes de ces acquis. Cette notion est encouragée par l'Organisation de coopération et de développement économiques (OCDE) dans le cadre du projet pilote intitulé Evaluation des résultats de l'enseignement supérieur (AHELO) (Tremblay *et al.*, 2012). Au sein de ce projet, l'OCDE a proposé une série d'acquis d'apprentissage complémentaires, au niveau international, afin d'aider les établissements d'enseignement supérieur à comparer les performances des étudiants (Tremblay *et al.*, 2012). La mesure des acquis peut être un indicateur prometteur d'enseignement efficace dans l'enseignement supérieur. Mais certains auteurs encouragent à la prudence dans ce domaine, pour plusieurs raisons.

Certains ont en effet constaté que la réussite relative des étudiants en ce qui concerne leurs acquis est causée par une combinaison de plusieurs facteurs, tels que «l'enseignement, la pratique et les compétences comportementales, ainsi que d'autres éléments extérieurs aux capacités de l'établissement» (Hénard, 2010, p. 84). La réussite relative des étudiants peut par conséquent donner une idée extrêmement fautive et erronée en tant qu'indicateur d'enseignement efficace. La relative contribution de l'enseignement dans la réussite des étudiants doit être isolée de ces autres facteurs pour permettre de juger de la qualité de l'instruction. Par conséquent, la réalisation brute d'un acquis n'est peut-être pas un indicateur fiable de l'efficacité de l'enseignement, ce qui rend discutable la validité des interprétations qui en seront tirées. D'autres auteurs notent aussi que l'utilisation des acquis des étudiants comme indicateurs d'enseignement efficace est problématique dans la mesure où on pourrait douter de la fiabilité et de la validité des méthodes d'évaluation utilisées pour apprécier la performance universitaire des étudiants (Lodge et Bonsanquet, 2014; Sadler, 2009). Si les évaluations utilisées pendant les cours pour mesurer la performance des étudiants à partir des acquis d'apprentissage ne sont pas fiables, valides et équitables, par extension les interprétations faites à partir de ces résultats d'évaluation poseront également problème. Certains auteurs vont jusqu'à affirmer que seules des expériences soigneusement conçues et randomisées permettent d'apprécier l'effet de l'enseignement sur l'apprentissage des étudiants (Bowman et Seifert, 2011).

Les techniques statistiques de pointe seraient sans doute capables de surmonter certains de ces problèmes. Ces dix dernières années, les modèles d'analyse de la valeur ajoutée ont gagné en popularité dans l'éducation postsecondaire (par exemple, OCDE, 2008; National Research Council, 2010). On s'est récemment penché sur l'utilité de ces modèles dans l'enseignement supérieur pour comparer l'efficacité des établissements en matière de soutien à l'apprentissage des étudiants (par exemple, Cunha et Miller, 2014; OCDE, 2013). Ils peuvent être utilisés pour fournir des estimations non biaisées de l'enseignement efficace au moyen d'un suivi des progrès universitaires sur plusieurs années. S'agissant de l'éducation postsecondaire, certains estiment que les modèles à valeur ajoutée «ont contribué à placer le débat sur la qualité de l'enseignant là où il doit être, à savoir l'amélioration des acquis des élèves comme première finalité de l'enseignement. Ils ont aussi apporté la promesse d'un élément quantitatif ô combien nécessaire pour l'évaluation de l'enseignement, tout en incitant au réexamen des questions d'équité et de bon usage des tests.» (Braun, 2005, p. 9) Un nombre croissant de recherches expérimentales ou quasi expérimentales au sein de l'éducation postsecondaire est venu donner un premier soutien à la possibilité, grâce à ces techniques de modélisation, d'estimer l'impact de l'enseignement sur l'amélioration des résultats des élèves aux tests (Chetty *et al.*, 2014; Kane et Staiger, 2008; Kane *et al.*, 2013). Malgré ces signes encourageants, dans la pratique, on se heurte toujours à ce problème essentiel, qui est que «si l'objectif est de trouver des causes, alors aucun modèle statistique, même le plus complexe, ni aucune méthode d'analyse, même la plus raffinée, ne pourront pleinement compenser l'absence de randomisation» (Braun, 2005, p. 8). De plus amples recherches «sont nécessaires pour comprendre si les estimations [basées sur l'analyse de la valeur ajoutée] devraient (ou ne devraient pas) être combinées à d'autres instruments de mesure pour repérer et retenir des enseignants efficaces» (Chetty *et al.*, 2014). Les enseignements tirés de l'éducation postsecondaire peuvent procurer d'importants éclairages sur l'utilisation de cette méthode pour évaluer l'enseignement dans l'enseignement supérieur, dont il sera utile de tenir compte.

Ce tour d'horizon des méthodes et des instruments de mesure actuellement utilisés pour évaluer l'enseignement dans l'enseignement supérieur suggère qu'une seule méthode ne suffit pas pour mesurer l'enseignement efficace. Les auteurs sont plutôt pour le recours à un éventail de méthodes (Hammonds *et al.*, 2017), notamment les plus récentes et innovantes (Abrami *et al.*, 2007; Hénard, 2010; Saroyan et Amundsen, 2001). Cela peut comprendre la triangulation de plusieurs types de données d'évaluation (Hénard, 2010; Hénard et Roseveare, 2012; Lodge et Bonsanquet, 2014; Saroyan et Amundsen, 2001; Spooen *et al.*, 2013). L'alignement de données d'évaluation puisées à différentes sources peut fournir un

tableau plus global et plus précis de l'enseignement que ne peut le faire aucune méthode isolée. La sélection stratégique de ces méthodes peut aussi présenter l'avantage de profiter des points forts de certaines pour contrebalancer les faiblesses de certaines autres. Avant toute chose, il faut que soient évaluées la fiabilité, la validité et l'équité des méthodes d'évaluation pour que les universités puissent donner le poids qui convient aux résultats (Uttl *et al.*, 2017). Ainsi, les méthodes qui ont fait l'objet d'une évaluation systématique et ont été jugées fiables, valides et équitables devraient avoir plus de poids que les méthodes aux résultats incertains.

Une formation spécialisée des personnels peut s'avérer nécessaire pour favoriser l'élaboration et la mise en œuvre des évaluations et l'analyse et l'interprétation de leurs résultats (Hénard et Roseveare, 2012; Linse, 2017; Spooren *et al.*, 2013). La formation des personnels chargés d'administrer et d'interpréter les évaluations de l'enseignement pourrait porter à la fois sur les méthodes qualitatives et quantitatives et la théorie de l'éducation. En outre, il pourrait être avantageux de fournir aux enseignants un accès aux spécialistes de ces questions, qui pourront les aider à interpréter et comprendre les résultats de leurs évaluations de l'enseignement.

Tendances et questions émergentes

Dans cette section finale, nous examinerons deux tendances et questions émergentes concernant l'évaluation de l'enseignement dans l'enseignement supérieur. La première concerne la précarisation du corps enseignant et la seconde évoque la question spécifique de l'apprentissage électronique et de l'enseignement en ligne. Nous examinerons les enjeux et les implications de ces nouvelles tendances pour l'évaluation de l'enseignement.

La précarisation du corps enseignant

La massification de l'enseignement supérieur a créé de nouvelles demandes en personnels dans les établissements d'enseignement supérieur. Cela a conduit partout dans le monde à une augmentation des emplois précaires, en tant qu'assistant, remplaçant, vacataire, contractuel ou enseignant non titularisable (Bryson, 2013; Coates et Goedegebuure, 2010; Parker, 2012; Harvey, 2017). Ces personnels sont le plus souvent employés sur une base horaire, contractuelle ou bénévole (Percy *et al.*, 2008). En Australie, par exemple, la majeure partie de l'enseignement universitaire est désormais assurée par des vacataires (May *et al.*, 2013). Aux Etats-Unis, l'enseignement s'appuie également sur des personnels non titularisables (Jaschik et Lederman, 2018). Le nombre des enseignants précaires est également en hausse au Canada, au Japon, en Nouvelle-Zélande et en Europe (Bryson, 2013; Sutherland et Gilbert, 2013; Sursock, 2015). Cette tendance devrait se poursuivre, et même augmenter (Jaschik et Lederman, 2018), dans la mesure où l'emploi d'enseignants sans poste est généralement plus rentable et qu'il permet de «libérer» des professeurs pour la recherche (Bryson, 2013). A titre d'exemple, dans le département des sciences sociales d'une université axée sur la recherche, au Royaume-Uni, la totalité des enseignements de premier cycle étaient assurés par des vacataires (Powney *et al.*, 2003). Il a été suggéré que cette augmentation des emplois précaires dans l'enseignement ne s'accompagne pas d'une intensification des approches pour garantir la qualité de l'apprentissage et de l'enseignement (Harvey, 2017).

Parce qu'ils changent constamment d'emploi, on dispose de peu de données concernant ces enseignants précaires (Bryson, 2013). Il est donc «plus difficile d'élaborer des approches systématiques pour reconnaître et soutenir ces personnels lorsqu'on ignore leur identité, leur localisation et leurs besoins en matière de développement professionnel» (Harvey, 2017, p. 1). La précarisation de l'enseignement dans les établissements d'enseignement supérieur crée donc de nouveaux défis pour l'évaluation, le suivi et le soutien méthodiques de la qualité

de l'enseignement lorsque celui-ci est assuré par des personnels précaires. Un problème majeur est l'inégalité des chances offertes à ces enseignants en matière de soutien et de traitement (Bryson, 2013). Ainsi, ils n'ont souvent pas accès au développement professionnel à des fins d'insertion, de mentorat, de préparation des cours et de formation (Anderson, 2007). D'autres auteurs constatent que les enseignants précaires ont souvent des difficultés pour accéder à des ressources essentielles pour leur enseignement comme les services de bibliothèque et les espaces de bureau (Kimber, 2003). Cet environnement peut créer des inégalités dans la possibilité d'exercer. Le manque d'accès aux ressources peut notamment être un obstacle à la réalisation d'un enseignement efficace. Cela peut avoir aussi d'importantes conséquences pour leur rétention et leur reconduction. Selon certains, les pressions causées par la nécessité d'être bien noté dans les évaluations de l'enseignement par les étudiants alors qu'ils n'ont pas accès aux ressources nécessaires peut avoir des conséquences négatives indésirables sur la qualité de l'apprentissage des étudiants. Plusieurs études suggèrent, à cet égard, que les enseignants précaires attribuent généralement des notes plus élevées que leurs collègues employés à plein temps (par exemple, Fedler *et al.*, 1989; Goldberg et Callahan, 1991; Salamonson *et al.*, 2010). Une des raisons possibles de ce décalage est que les personnels précaires peuvent se montrer plus indulgents afin d'obtenir de meilleures évaluations de leur enseignement par les étudiants (Moore et Trahan, 1998; Salamonson *et al.*, 2010). Par conséquent, le manque de formation et de ressources combiné aux pressions à la performance pourraient avoir des conséquences fâcheuses sur la qualité de l'enseignement et de l'apprentissage. C'est pourquoi un examen plus approfondi de la manière dont ces enseignants précaires sont évalués et soutenus s'impose. Certains suggèrent que «la collecte de données complètes et précises», concernant notamment l'efficacité de l'enseignement, peut permettre de signaler les besoins en matière de développement professionnel (Harvey, 2017).

L'apprentissage électronique et l'enseignement en ligne

Une deuxième tendance émergente est le développement des cours en ligne et de l'éducation à distance. Depuis dix ans, le nombre d'étudiants qui se forment via Internet a augmenté de plusieurs millions (Allen et Seaman, 2015). De nombreux établissements d'enseignement supérieur inscrivent l'apprentissage en ligne parmi leurs stratégies à long terme (Allen et Seaman, 2011; He *et al.*, 2014). La proportion des contenus proposés en ligne peut être très variable selon le type de programme suivi. Selon la définition d'Allen et Seaman (2015), les cours soutenus par le Web dispensent 1 à 29 pour cent de leurs contenus en ligne et se déroulent généralement en présentiel, en s'appuyant sur des plates-formes d'apprentissage en ligne (LMS) ou sur des pages Web pour publier le contenu du programme et des devoirs. Les cours mixtes ou hybrides combinent quant à eux instruction en ligne et enseignement en présentiel, mais avec une part substantielle de contenu assisté en ligne (de 30 à 79 pour cent), comme les débats en ligne, et une part moindre de présence directe. Les cours en ligne ne prévoient par contre aucun face-à-face direct, leur contenu étant intégralement diffusé en ligne.

Comme il s'agit d'un domaine nouveau pour la science de l'enseignement et de l'apprentissage, bien des questions restent en suspens et exigent d'être approfondies. Nous en évoquerons brièvement trois. La première concerne la mise en œuvre d'un enseignement en ligne efficace. On ne sait pas actuellement comment définir l'enseignement efficace dans les environnements en ligne ni en quoi il serait différent de l'enseignement dispensé dans un environnement d'apprentissage en présentiel. Ce premier problème a deux implications pour l'évaluation de l'enseignement dans l'enseignement supérieur. La première est que les étudiants pourraient avoir une appréciation différente de l'enseignement en ligne par rapport à l'enseignement dispensé dans les environnements d'apprentissage traditionnels (Young, 2006, p. 65). La seconde est qu'il faut réunir des facteurs particuliers pour assurer un enseignement efficace lorsqu'il est dispensé en ligne (Sun *et al.*, 2008). Certains suggèrent, par exemple, que l'enseignement efficace dans les environnements d'apprentissage

électroniques requiert des compétences techniques, des compétences de modération en ligne et des compétences de communication en ligne (Gunn, 2001). Cela signifie que les évaluations de l'enseignement qui ont actuellement cours doivent être ajustées ou reconçues pour mesurer des indicateurs pertinents qui permettent des interprétations valides de l'enseignement efficace dans ce type d'environnements. La raison en est que «la conception et la mise en œuvre de stratégies d'évaluation de l'enseignement en ligne se distinguent nécessairement de celles concernant l'environnement en présentiel; elles varient en fonction de la nature, de la finalité et de l'objet de l'évaluation» (Tobin *et al.*, 2015, p. 15). Ainsi, l'évaluation, par les étudiants, d'outils pédagogiques conçus pour l'enseignement en présentiel ne contient pas de questions permettant d'évaluer l'enseignement de qualité lorsque l'instruction se fait en ligne (Bangert, 2006). Il n'apparaît pas non plus clairement si d'autres méthodes d'évaluation de l'enseignement utilisées dans le cadre d'une instruction en présentiel pourraient être adaptées aux environnements en ligne. On ne sait pas s'il serait possible d'utiliser les observations de l'enseignement dans les situations d'enseignement en ligne. Par conséquent, les tentatives pour appliquer les méthodes d'évaluation existantes aux environnements d'enseignement en ligne sont sans doute injustifiées. Un second problème est qu'il est difficile de séparer l'enseignement des éléments de conception de la plate-forme d'apprentissage en ligne (Tobin *et al.*, 2015, p. 15). Les évaluations de l'enseignement dispensé dans des environnements en ligne devraient être conçues de façon à ne mesurer que les dimensions relevant des compétences de l'enseignant.

Mais l'enseignement en ligne offre aussi de nouvelles possibilités d'évaluation qui n'existent pas dans les environnements d'apprentissage en présentiel. Les plates-formes d'apprentissage en ligne peuvent collecter des données de trace concernant l'instructeur qui pourraient servir d'indicateurs d'enseignement efficace. Parmi les indicateurs à considérer, citons la fréquence des connexions, le temps passé sur les LMS, la nature et la qualité des commentaires en retour, la rapidité des notations et annotations et la proportion des posts transmis par l'instructeur à l'étudiant (Tobin *et al.*, 2015, p. 15). Il s'agit de nouvelles possibilités, mais on manque encore d'éléments de preuve pour justifier l'utilisation de ces indicateurs. Bien que les LMS puissent apporter de nouvelles sources de données sur les interactions des instructeurs, il est important de séparer celles qui présentent un intérêt pour juger de l'efficacité de l'enseignement de celles qui n'en ont pas.

Résumé des recommandations

L'examen des normes de mesures et d'évaluation en vigueur et des méthodes et instruments d'évaluation de l'enseignement dans l'enseignement supérieur suggère plusieurs recommandations politiques. Elles sont résumées ci-dessous:

- La tendance en faveur de l'assurance-qualité dans l'enseignement supérieur pose aux établissements de nouveaux défis en matière d'élaboration et de réalisation d'un enseignement efficace. Toutefois, la notion d'efficacité souffre d'un excès de conceptualisation, et les évaluations de l'enseignement manquent largement de cohérence d'un établissement, d'un contexte, d'une discipline et d'un pays à l'autre. On pourrait donc concevoir des politiques destinées à promouvoir l'adoption de normes cohérentes d'enseignement efficace et d'évaluation de la qualité, en encourageant notamment le recours à un éventail de méthodes d'évaluation, de façon à appréhender les multiples dimensions de l'enseignement efficace.
- L'assurance-qualité appelle aussi à des politiques capables d'indiquer le bon usage des principes de mesure et d'évaluation de l'enseignement. Ces politiques pourraient donner des directives concernant l'application des principes de fiabilité, de validité et d'équité dans la conception et la réalisation des évaluations et l'interprétation de leurs résultats. Elles pourront aussi aider à définir des normes pour la collecte de données attestant que les pratiques d'évaluation sont fiables et permettent des interprétations

valides et équitables des résultats obtenus. Ces attestations sont nécessaires pour nourrir la confiance dans les interprétations concernant l'enseignement, qui ont d'importantes ramifications en matière de décision concernant les personnels et de suivi et de poursuite d'une éducation de qualité. Il convient également de tenir compte des tendances émergentes, notamment en ce qui concerne le perfectionnement des personnels d'enseignement précaires et les contextes d'enseignement en ligne.

- Dans le même ordre d'idées, des orientations sont nécessaires concernant la réponse que les établissements devraient apporter en cas de non-fiabilité des évaluations ou d'interprétations inexacts ou inéquitables de l'enseignement. Lorsqu'il est constaté, par exemple, que l'évaluation d'un instrument d'enseignement par les étudiants biaise de façon systématique et négative les résultats concernant une enseignante, par exemple, ces résultats devraient être corrigés statistiquement ou adéquatement pondérés et l'outil d'évaluation ajusté pour ses usages futurs.
- Des politiques devraient être élaborées et diffusées auprès des personnels enseignants pour encourager à la transparence dans les normes définies d'enseignement efficace, les modalités de mesure et l'utilisation des résultats au sein des établissements. Les établissements devraient, par exemple, être encouragés à prévoir des politiques d'évaluation interne. Des politiques internes transparentes peuvent contribuer à l'adoption de pratiques d'évaluation cohérentes au sein des établissements ainsi qu'au contrôle et au maintien d'une éducation de qualité. Ces politiques peuvent aussi informer les enseignants de ce qu'on attend d'eux pour qu'ils soient reconnus comme des enseignants efficaces.
- Il convient aussi d'aligner les politiques institutionnelles pour indiquer que l'enseignement est valorisé, le choix des indicateurs d'enseignement efficace, les outils d'évaluation et les incitations.
- Les politiques devraient encourager les établissements à fournir des possibilités de formation et de développement professionnel afin de favoriser des pratiques d'évaluation de qualité. Avec les possibilités d'adoption de méthodes statistiques d'évaluation plus sophistiquées par les établissements, comme les modèles à valeur ajoutée, la formation revêtira une importance d'autant plus grande. Des services de formation devraient donc être fournis aux personnels chargés d'élaborer, d'administrer et d'interpréter les évaluations. Des formations sont nécessaires, par exemple, sur les théories de mesure et d'évaluation de l'éducation, les méthodes d'évaluation, la bonne maîtrise de l'élaboration, de l'administration et de l'interprétation (y compris des principes de fiabilité, de validité et d'équité), et les analyses qualitatives et quantitatives. Il peut également être avantageux de fournir aux enseignants un accès aux spécialistes des mesures et des évaluations qui peuvent les aider à interpréter et à comprendre les résultats de leurs évaluations de l'enseignement. En outre, des spécialistes crédibles devraient être disponibles au sein des universités pour fournir un développement professionnel aux personnels d'enseignement en ce qui concerne l'enseignement.

Domaines de recherche future

Les données internationales concernant l'évaluation de l'enseignement dans l'enseignement supérieur à des fins de perfectionnement et de reddition des comptes sont absentes ou difficiles à obtenir. Des recherches supplémentaires sont nécessaires dans plusieurs domaines si l'on veut fournir des recommandations de bonne pratique appuyées sur des données probantes. Ces domaines de recherche future sont résumés ci-dessous:

-
- la mise au point de cadres ou de normes universellement reconnus pour définir et évaluer l'enseignement efficace, pouvant contribuer à assurer la qualité de l'enseignement supérieur dans le monde; compte tenu de la complexité de l'enseignement dans l'enseignement supérieur, ces normes exigent des recherches supplémentaires;
 - un examen de la manière dont les établissements recueillent, pondèrent et utilisent les différentes méthodes d'évaluation de l'enseignement dans l'enseignement supérieur: l'objectif serait de fournir un premier étalonnage des pratiques d'évaluation au sein des établissements, qui engloberait les politiques internes, les méthodes et les instruments de mesure, les preuves de la qualité de l'évaluation, les ressources et la formation, et l'utilisation des résultats;
 - la continuation de l'examen des méthodes d'évaluation émergentes, y compris les modèles à valeur ajoutée et les données de trace de l'enseignement en ligne dans le cadre de l'enseignement supérieur: cette recherche pourrait nous éclairer sur l'utilité de ces méthodes et servir à l'élaboration de nouvelles directives et politiques;
 - une compilation de données internationales sur les personnels précaires, portant notamment sur leurs besoins en matière de développement professionnel.

Conclusions

La massification de l'enseignement supérieur, les progrès de l'assurance-qualité, les changements dans la composition du corps enseignant, l'essor de l'enseignement en ligne et les besoins accrus de développement professionnel ont créé de nouvelles demandes et de nouveaux défis en matière d'évaluation de l'enseignement dans l'enseignement supérieur. Ils ont surtout mis en évidence la nécessité de créer des systèmes d'évaluation de la qualité qui permettent aux établissements d'évaluer l'efficacité de l'enseignement à des fins de perfectionnement et de reddition des comptes. Cette courte étude a montré la complexité de l'élaboration de tels systèmes, mais aussi les possibilités de faire progresser les pratiques d'évaluation dans l'enseignement supérieur. L'évaluation de l'enseignement doit s'appuyer sur des conceptualisations précises de l'enseignement efficace afin de définir les normes et les attentes de l'enseignement. On ne dispose pas encore actuellement de normes mondiales d'enseignement efficace, mais cet objectif peut être atteint en s'appuyant sur les définitions locales existantes. De plus, la conception, l'administration et l'interprétation des évaluations de l'enseignement doivent obéir à des normes de mesure reconnues. Cela signifie que les personnels chargés d'évaluer l'enseignement doivent avoir accès aux possibilités de formation nécessaires pour garantir la qualité des systèmes d'évaluation des établissements. Bien que le présent examen reste limité par la disponibilité des données internationales en la matière, il constitue une bonne base d'informations pour juger de l'état actuel des évaluations de l'enseignement dans l'enseignement supérieur et des domaines encore lacunaires. Pour dispenser une éducation de qualité dans l'enseignement supérieur, il est indispensable de promouvoir des systèmes d'évaluation de l'enseignement qui soient de qualité et basés sur des données probantes.

Bibliographie

- Abrami, P.C. (2001). «Improving judgments about teaching effectiveness using teacher rating forms», dans *New Directions for Institutional Research*, 27(5), pp. 59-87.
- Abrami, P.C., d'Apollonia, S. et Rosenfield, S. (2007). «The dimensionality of student ratings of instruction: What we know and what we do not», dans *The scholarship of teaching and learning in higher education: An evidence-based perspective*, pp. 385-456: Springer.
- Abrami, P.C., Dickens, W.J., Perry, R.P. et Leventhal, L. (1980). «Do teacher standards for assigning grades affect student evaluations of instruction?», dans *Journal of Educational Psychology*, 72(1), pp. 107-118.
- Allen, I.E. et Seaman, J. (2011). *Going the distance: Online education in the United States*, 2011. Sloan Consortium, PO Box 1238, Newburyport, MA 01950.
- . (2015). *Grade Level: Tracking Online Education in the United States*. Babson Survey Research Group.
- Altbach, P., Reisberg, L. et Rumbley, L. (2009). *Trends in Global Higher Education: Tracking an Academic Revolution*, rapport préparé pour la Conférence mondiale sur l'enseignement supérieur de l'UNESCO, 2009 (Paris, UNESCO).
- American Educational Research Association, American Psychological Association and National Council on Measurement in Education (2014). *Standards for educational and psychological testing*.
- Anderson, V. (2007). «Contingent and marginalised? Academic development and part-time teachers», dans *International Journal for Academic Development*, 12(2), pp. 111-121.
- Bangert, A.W. (2006). «The development of an instrument for assessing online teaching effectiveness», dans *Journal of Educational Computing Research*, 35(3), p. 227.
- Banque mondiale: [Statistiques sur l'éducation de la Banque mondiale](#).
- Berk, R.A. (2005). «Survey of 12 strategies to measure teaching effectiveness», dans *International Journal of Teaching and Learning in Higher Education*, 17(1), pp. 48-62.
- . 2013. *Top 10 flashpoints in student ratings and the evaluation of teaching: What faculty and administrators must know to protect themselves in employment decisions*. Stylus Publishing, LLC.
- Biggs, J. (2001). «The reflective institution: Assuring and enhancing the quality of teaching and learning», dans *Higher education*, 41(3), pp. 221-238.
- Bowman, N.A. et Seifert, T.A. (2011). «Can college students accurately assess what affects their learning and development?», dans *Journal of College Student Development*, 52(3), pp. 270-290.
- Braskamp, L.A. et Ory, J.C. (1994). *Assessing Faculty Work: Enhancing Individual and Institutional Performance*. Jossey-Bass Higher and Adult Education Series (San Francisco, CA, Jossey-Bass Inc.).
- Braun, H. I. (2005). «Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models. Policy Information Perspective» dans *Educational Testing Service*.

-
- Bryson, C. (2013). *Supporting sessional teaching staff in the UK – to what extent is there real progress?*, recherche en ligne.
- Centra, J.A. (1993). *Reflective Faculty Evaluation: Enhancing Teaching and Determining Faculty Effectiveness*, Jossey-Bass Higher and Adult Education Series (San Francisco, CA, Jossey-Bass Inc.).
- Chetty, R., Friedman, J.N. et Rockoff, J.E. (2014). «Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates», dans *American Economic Review*, 104(9), pp. 2593-2632.
- Chetty, R., Friedman, J.N. et Rockoff, J.E. (2014). «Discussion of the American statistical association's statement (2014) on using value-added models for educational assessment», dans *Statistics and Public Policy*, 1(1), pp. 111-113.
- Chickering, A.W. et Gamson, Z.F. (1987). «Seven principles for good practice in undergraduate education», dans *American Association for Higher Education Bulletin*, pp. 3-7.
- Chism, N. V. N. (1999). *Peer Review of Teaching. A Sourcebook* (Bolton, MA: Anker).
- Chism, N.V.N. et Chism, G.W. (2007). *Peer review of teaching: A sourcebook* (Bolton, MA: Anker Pub. Co.).
- Clayson, D.E. (2013). «[Initial Impressions and the Student Evaluation of Teaching](#)», dans *Journal of Education for Business*, 88(1), pp. 26-35.
- Clayson, D.E. et Haley, D.A. (2011). «[Are students telling us the truth? A critical look at the student evaluation of teaching](#)», dans *Marketing Education Review*, 21(2), pp. 101-112.
- Coates, H. et Goedegebuure, L. (2010). *The real academic revolution*, présentation de recherche (Carlton: [LH Martin Institute](#)).
- Coffey, M. et Gibbs, G. (2001). «The evaluation of the student evaluation of educational quality questionnaire (SEEQ) in UK higher education», dans *Assessment & Evaluation in Higher Education*, 26(1), pp. 89-93.
- Cosh, J. (1998). «Peer observation in higher education – A reflective approach», dans *Innovations in Education and Training International*, 35(2), pp. 171-176.
- Cunha, J. M. et Miller, T. (2014). «Measuring value-added in higher education: Possibilities and limitations in the use of administrative data» dans *Economics of Education Review*, 42, pp. 64-77.
- d'Apollonia, S. et Abrami, P.C. (1997). «Navigating student ratings of instruction», dans *American Psychologist*, 52(11), pp. 1198-1208.
- Darwin, S. (2017). «[What contemporary work are student ratings actually doing in higher education?](#)», dans *Studies in Educational Evaluation*, 54, pp. 13-21.
- DeZure, D. (1999). «Evaluating teaching through peer classroom observation», dans P. Seldin (dir. de publication) *Changing Practices in Evaluating Teaching: A Practical Guide to Improved Faculty Performance and Promotion/Tenure Decisions* (Bolton, MA: Anker).
- Eiszler, C.F. (2002). «College students' evaluations of teaching and grade inflation», dans *Research in Higher Education*, 43(4), pp. 483-501.

-
- Fedler, F., Counts, T., Stoner, K.R. (1989). «Adjunct profs grade higher than faculty at three schools», dans *Journalism Educator*, 44, pp. 32-37.
- Feldman, K.A. (1976). «Grades and college students' evaluations of their courses and teachers», dans *Research in Higher Education. Journal of the Association for Institutional Research*, 4(1), pp. 69-111.
- Freng, S. et Webber, D. (2009). «Turning up the heat on online teaching evaluations: Does "hotness" matter?», dans *Teaching of Psychology*, 36(3), pp. 189-193.
- Goldberg, G. et Callahan, J. (1991). «Objectivity of student evaluations of instructors», dans *Journal of Education for Business*, 66, pp. 377-378.
- Gosling, D. (2009). «A new approach to peer review of teaching», dans D. Gosling et K. Mason O'Connor (dir. de publication): *Beyond the peer observation of teaching*, pp. 7-15. SEDA Paper 124 (Londres, Staff and Educational Development Association Ltd.).
- Greenwald, A.G. et Gillmore, G.M. (1997). «Grading leniency is a removable contaminant of student ratings», dans *American psychologist*, 52(11), p. 1209.
- Gunn, C. (2001). «Effective online teaching: How far do the frameworks go», dans *Meeting at the Crossroads: Proceedings of the Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCLITE)*, pp. 235-244.
- Hammonds, F., Mariano, G.J., Ammons, G. et Chambers, S. (2017). «Student evaluations of teaching: improving teaching quality in higher education», dans *Perspectives: Policy and Practice in Higher Education*, 21(1), p. 26.
- Harvey, L. (2018). [Analytic Quality Glossary](#), Quality Research International.
- Harvey, M. (2017). *Quality learning and teaching with sessional staff: systematising good practice for academic development*. Taylor & Francis.
- He, W., Xu, G. et Kruck, S. (2014). «Online IS education for the 21st century», dans *Journal of Information Systems Education*, 25(2), p. 101.
- Hénard, F. (2010). *Learning our lesson: Review of quality teaching in higher education* (vol. 2010, n° 2), résumé: [Retenir la leçon. Enquête sur l'enseignement de qualité dans l'enseignement supérieur](#) (Paris, OCDE).
- Hénard, F. et Roseveare, D. (2012). «Fostering quality teaching in higher education: Policies and Practices» dans *An IMHE Guide for Higher Education Institutions*, pp. 7-11.
- Hobson, S. M. et Talbot, D. M. (2001). «Understanding student evaluations: What all faculty should know» dans *College teaching*, 49(1), pp. 26-31.
- Jaschik, S. et Lederman, D. (2018). [The 2018 Inside Higher Ed Survey of College & University Chief Academic Officers, Inside Higher Ed](#).
- Jones, S.J. (2012). «Reading between the lines of online course evaluations: Identifiable actions that improve student perceptions of teaching effectiveness and course value», dans *Journal of Asynchronous Learning Networks*, 16(1), pp. 49-58.
- Kane, T.J. et Staiger, D.O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation*, n° 14607, National Bureau of Economic Research.

-
- Kane, T.J., McCaffrey, D.F., Miller, T. et Staiger, D.O. (2013). «Have we identified effective teachers? Validating measures of effective teaching using random assignment», dans *Research Paper*, MET Project, Bill & Melinda Gates Foundation.
- Kember, D., Ma, R. et McNaught, C. (2006). *Excellent university teaching* (Hong-kong, Chine, The Chinese University Press).
- Kimber, M. (2003). «The tenured ‘core’ and the tenuous ‘periphery’: the casualisation of academic work in Australian universities» dans *Journal of Higher Education Policy and Management*, 25(1), pp. 41-50.
- Kinchin, I. M. (2005). «Evolving diversity within a model of peer observation at a UK university», dans *British Educational Research Association (BERA) Annual Conference, University of Glamorgan (Wales)*, pp. 14-17.
- Lewandowski, G.W. Jr, Higgins, E. et Nardone, N.N. (2011). «Just a harmless website?: An experimental examination of RateMyProfessors.com’s effect on student evaluations», dans *Assessment & Evaluation in Higher Education*, 37(8), pp. 987-1002.
- Linse, A.R. (2017). «Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees», dans *Studies in Educational Evaluation*, 54, pp. 94-106.
- Lodge, J.M. et Bonsanquet, A. (2014). «Evaluating quality learning in higher education: re-examining the evidence», dans *Quality in Higher Education*, 20(1), pp. 3-23.
- MacNell, L., Driscoll, A. et Hunt, A.N. (2015). «What’s in a name: Exposing gender bias in student ratings of teaching», dans *Innovative Higher Education*, 40(4), pp. 291-303.
- Marsh, H.W. (2007). «Students’ evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness», dans *The scholarship of teaching and learning in higher education: An evidence-based perspective*, pp. 319-383: Springer.
- Marsh, H.W. et Roche, L.A. (1997). «Making students’ evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility», dans *American Psychologist*, 52(11), pp. 1187-1197.
- May, R., Strachan, G. et Peetz, D. (2013). «Workforce Development and Renewal in Australian Universities and the Management of Casual Academic Staff» dans *Journal of University Teaching and Learning Practice*, 10(3), p. 3.
- McKeachie, W.J. (1979). «Student ratings of faculty: A reprise», dans *Academe*, 65(6), pp. 384-397.
- McMahon, T., Barret, T. et O’Neill, G. (2007). «Using observation of teaching to improve quality: Finding your way through the muddle of competing conceptions, confusion of practice and mutually exclusive intentions», dans *Teaching in Higher Education*, 12(4), pp. 499-511.
- Ministère des Entreprises, de l’Innovation et des Compétences du Royaume-Uni (2015). *Fulfilling our potential: Teaching excellence, social mobility and student choice*, Department for Business, Innovation and Skills (Londres, BIS).
- Moore, M. et Trahan, R. (1998). «Tenure status and grading practice», dans *Sociological Perspectives*, 41(4), pp. 775-782.

-
- Muijs, D. (2006). «Measuring teacher effectiveness: Some methodological reflections», dans *Educational research and evaluation*, 12(1), pp. 53-74.
- Nargundkar, S. et Shrikhande, M. (2014). «Norming of student evaluations of instruction: Impact of non-instructional factors», dans *Decision Sciences: Journal of Innovative Education*, 12(1), pp. 55-72.
- Nasser-Abu Alhija, F. (2017). «Teaching in higher education: Good teaching through students' lens», dans *Studies in Educational Evaluation*, 54, pp. 4-12.
- National Research Council (2010). *Getting value out of value-added*, H. Braun, N. Chudowsky et J. Koenig (dir. de publication) (Washington, DC: National Academies Press).
- OCDE (Organisation de coopération et de développement économiques) (2008). *Measuring improvements in learning outcomes: Best practices to assess the value-added of schools* (Paris, OCDE).
- . (2009). *Teacher Evaluation: A Conceptual Framework and examples of Country Practices* (Paris, OCDE).
- . (2013). *Assessment of Higher Education Learning Outcomes: Feasibility Study Report* (vol. 3: Further Insights) (Paris, OCDE).
- . (2017). *Enhancing Higher Education System Performance* (Paris, OCDE).
- Onwuegbuzie, A.J., Witcher, A.E., Collins, K.M.T., Filer, J.D., Wiedmaier, C.D. et Moore, C.W. (2007). «Students perceptions of characteristics of effective college teachers: A validity study of a teaching evaluation form using a mixed-methods analysis», dans *American Educational Research Journal*, 44(1), pp. 113-160.
- Parker, L.D. (2012). «From privatised to hybrid corporatised higher education: A global financial management discourse», dans *Financial Accountability and Management*, 28(3), pp. 247-268.
- Paulsen, M.B. (2002). «Evaluating teaching performance», dans *New Directions for Institutional Research*, 2002 (114), pp. 5-18.
- Percy, A., Scoufis, M., Parry, S., Goody A., Hicks M., Macdonald I., Martinez K. *et al.* (2008). *The RED Report: Recognition – Enhancement – Development: The contribution of sessional teachers to higher education*. Sydney: Australian Learning and Teaching Council.
- Pitman, T. (2014). «Reinterpreting higher education quality in response to policies of mass education: The Australian experience», dans *Quality in Higher Education*, 20(3), pp. 348-363.
- Pounder, J.S., Ho Hung-lam, E. et Groves, J.M. (2016). «Faculty-student engagement in teaching observation and assessment: A Hong Kong initiative», dans *Assessment & Evaluation in Higher Education*, 41(8), p. 1193.
- Powney, J., Bryson, C., Cloonan, M. et Wilson, V. (2003). *The Appointment, Retention and Promotion of Academic Staff within Higher Education Institutions* (Bristol, HEFCE).
- Prosser, M. et Trigwell, K. (1997). «Relations between perceptions of the teaching environment and approaches to teaching», dans *British Journal of Educational Psychology*, 67(1), pp. 25-35.

-
- Ramsden, P. (2003). «Learning to teach in higher education» (Londres, Routledge).
- Remedios, R. et Lieberman, D.A. (2008). «I liked your course because you taught me well: The influence of grades, workload, expectations and goals on students' evaluations of teaching», dans *British Educational Research Journal*, 34(1), pp. 91-115.
- Sadler, D.R. (2009). «Indeterminacy in the use of preset criteria for assessment and grading», dans *Assessment & Evaluation in Higher Education*, 34(2), pp. 159-179.
- Salamonson, Y., Halcomb, E.J., Andrew, S., Peters, K. et Jackson, D. (2010). «A comparative study of assessment grading and nursing students' perceptions of quality in sessional and tenured teachers», dans *Journal of Nursing Scholarship*, 42(4), pp. 423-429.
- Salmi, J. et Saroyan, A. (2007). «League tables as policy instruments», dans *Higher Education Management and Policy*, 19(2), pp. 1-38.
- Santiago, P., Tremblay, K., Basri, E. et Arnal, E. (2008). *Tertiary education for the knowledge society* (vol. 1) (Paris, OCDE).
- Saroyan, A. et Amundsen, C. (2001). «Evaluating University Teaching: Time to take stock», dans *Assessment & Evaluation in Higher Education*, 26(4), pp. 341-353.
- Saroyan, A., Amundsen, C., McAlpine, L., Weston, C., Winer, L. et Gandell, T. (2004). «Assumptions underlying workshop activities», dans A. Saroyan et C. Amundsen (dir. de publication): *Rethinking teaching in higher education*, pp. 15-29. Sterling, VA: Stylus.
- Seldin, P. (1993). *Successful use of teaching portfolios* (Bolton, MA: Anker Pub. Co.).
- Simpson, P.M. et Sigauw, J.A. (2000). «Student evaluations of teaching: An exploratory study of the faculty response», dans *Journal of Marketing Education*, 22(3), pp. 199-213.
- Skelton, A. (2005). *Understanding teaching excellence in higher education: Towards a critical approach* (Londres: Routledge).
- Smith, H. (2012). «The Unintended Consequences of Grading Teaching» dans *Teaching in higher education*, 17(6), p. 747.
- Spooren, P., Brockx, B. et Mortelmans, D. (2013). «On the validity of student evaluation of teaching: The state of the art», dans *Review of Educational Research*, 83(4), pp. 598-642.
- Stanfield, D.A. et Shimmi, Y. (2014). «Chinese higher education: Statistics and trends», dans L.E. Rumbley, R.M. Helms, P. McGill Peterson et P.G. Altbach (dir. de publication): *Global Opportunities and Challenges for Higher Education Leaders*, pp. 77-82 (Rotterdam, Sense Publishers).
- Stein, S.J., Spiller, D., Terry, S., Harris, T., Deaker, L. et Kennedy, J. (2012). «Unlocking the impact of tertiary teachers' perceptions of student evaluation of teaching» (Wellington, Nouvelle-Zélande, Ako Aotearoa National Centre for Tertiary Teaching Excellence).
- Stowell, J.R., Addison, W.E. et Smith, J.L. (2012). «Comparison of online and classroom-based student evaluations of instruction», dans *Assessment & Evaluation in Higher Education*, 37(4), pp. 465-473.

-
- Strang, L., Bélanger, J., Manville, C., Meads, C. (2016). *Review of the Research Literature on Defining and Demonstrating Quality Teaching and Impact in Higher Education*, [Higher Education Academy](#) (York, Royaume-Uni).
- Stumpf, S.A. et Freedman, R.D. (1979). «Expected grade covariation with student ratings of instruction: Individual versus class effects», dans *Journal of Educational Psychology*, 71(3), pp. 293-302.
- Sun, P.-C., Tsai, R.J., Finger, G., Chen, Y.-Y. et Yeh, D. (2008). «What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction», dans *Computers and Education*, 50(4), pp. 1183-1202.
- Surgenor, P.W.G. (2013). «Obstacles and opportunities: Addressing the growing pains of summative student evaluation of teaching», dans *Assessment & Evaluation in Higher Education*, 38, pp. 363-376.
- Sursock, A. (2015). «Trends 2015: Learning and teaching in European universities» dans *Brussels: European University Association*.
- Sutherland, K.A. et Gilbert, A. (2013). «Academic aspirations amongst sessional tutors in a New Zealand University» dans *Journal of University Teaching & Learning Practice*, 10(3), p. 7.
- Tobin, T.J., Mandernach, B.J. et Taylor, A.H. (2015). *Evaluating online teaching: Implementing best practices*, John Wiley & Sons.
- Torres, A.C., Lopes, A., Valente, J.M.S. et Mouraz, A. (2017). «What catches the eye in class observation? Observers' perspectives in a multidisciplinary peer observation of teaching program», dans *Teaching in Higher Education*, 22(7), p. 822.
- Tremblay, K., Lalancette, D. et Roseveare, D. (2012). *Assessment of Higher Education Learning Outcomes: Feasibility Study Report*, vol. 1 (Paris, OCDE).
- Trigwell, K. et Prosser, M. (1996). «Changing approaches to teaching: A relational perspective», dans *Studies in Higher Education*, 21(3), pp. 275-284.
- Uttl, B., White, C.A. et Gonzalez, D.W. (2017). «[Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related](#)», dans *Studies in Educational Evaluation*, 54, pp. 22-42.
- Vincent-Lancrin, S. (2008). «What is the impact of demography on higher education systems? A forward-looking approach for OECD countries», dans *Higher education to 2030* (vol. 1), pp. 41-103.
- Widmeyer, W.N. et Loy, J.W. (1988). «When you're hot, you're hot! Warm-cold effects in first impressions of persons and teaching effectiveness», dans *Journal of Educational Psychology*, 80(1), p. 118.
- Winer, L., Di Genova, L., Vungoc, P.-A. et Talsma, S. (2012). *Interpreting end-of-course evaluation results* (Montréal: Teaching and Learning Services, McGill University).
- Young, S. (2006). «Student Views of Effective Online Teaching in Higher Education» dans *American Journal of Distance Education*, 20(2), p. 65.
- Zhao, J. et Gallant, D. J. (2012). «Student evaluation of instruction in higher education: Exploring issues of validity and reliability» dans *Assessment & Evaluation in Higher Education*, 37(2), pp. 227-235.