Mesure et évaluation en éducation

L'évaluation de l'enseignement universitaire par les étudiants : quelques pistes à suivre pour un meilleur usage

Hélène Poissant

Volume 17, numéro 3, 1995

URI : https://id.erudit.org/iderudit/1092315ar
DOI : https://doi.org/10.7202/1092315ar

Aller au sommaire du numéro

Éditeur(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (imprimé)
2368-2000 (numérique)

Découvrir la revue

Citer cet article

Poissant, H. (1995). L'évaluation de l'enseignement universitaire par les étudiants : quelques pistes à suivre pour un meilleur usage. *Mesure et évaluation en éducation*, *17*(3), 89–123. https://doi.org/10.7202/1092315ar

Résumé de l'article

Dans cet article, nous relevons certains problèmes liés à l'évaluation de l'enseignement universitaire par les étudiants et nous proposons quelques moyens pour y remédier. Différents points émergent et plusieurs sources et moyens d'information doivent être utilisés. Dans un contexte d'évaluation sommative, les items généraux des questionnaires sont préférables aux items spécifiques ; les différentes personnes en cause dans l'évaluation, professeurs, administrateurs et étudiants, doivent connaître les buts et les usages des résultats de l'évaluation.

Mesure et évaluation en éducation
Vol. 17 n° 3, pages 89 à 123.

L'évaluation de l'enseignement universitaire par les étudiants: quelques pistes à suivre pour un meilleur usage

Hélène Poissant
Teachers College
Columbia University

Dans cet article, nous relevons certains problèmes liés à l'évaluation de l'enseignement universitaire par les étudiants et nous proposons quelques moyens pour y remédier. Différents points émergent et plusieurs sources et moyens d'information doivent être utilisés. Dans un contexte d'évaluation sommative, les items généraux des questionnaires sont préférables aux items spécifiques; les différentes personnes en cause dans l'évaluation, professeurs, administrateurs et étudiants, doivent connaître les buts et les usages des résultats de l'évaluation.

évaluation de l'enseignement, enseignement universitaire, évaluation de cours, évaluation par les étudiants

---

This study examines some problems related with students' evaluation of university professors and suggests some answers. The major conclusions are: several sources and means should be used to get a complete information; when administrative decisions are taken, general items are preferable; all individuals concerned with evaluation (professors, administrators and students) should be better informed of the evaluation's goals and uses.

teacher evaluation, higher education, course evaluation, evaluation by students

Introduction

La problématique de l'évaluation de l'enseignement se pose de plus en plus au Québec et ailleurs. Nous comptons, par cet article, contribuer à conférer une qualité et une crédibilité accrues aux pratiques dans ce domaine en montrant comment éviter certains pièges dans lesquels les différentes personnes reliées à l'évaluation peuvent tomber en demandant aux étudiants de fournir une appréciation des cours.

L'évaluation de l'enseignement par les étudiants est devenue une pratique courante dans les universités canadiennes. Elle devient aussi

présentement de plus en plus généralisée mais aussi discutée dans le milieu collégial. En dépit des controverses qui l'entourent, on peut donc dire qu'elle est entrée, bon gré, mal gré, dans les *mœurs et coutumes* des milieux académiques. Dans le présent article, nous nous limiterons aux systèmes d'évaluation existant à l'université puisque, pour le moment encore, c'est à cet endroit qu'ils sont le plus largement implantés. De même, leur emploi fréquent dans les prises de décision affectant la promotion et la poursuite des contrats, des professeurs d'universités exige l'établissement de nouveaux critères de qualité hautement élevés. L'auteur de ce texte souscrit à l'idée que, trop souvent encore, des outils employés pour la sélection du personnel universitaire ne sont pas appropriés à un tel contexte d'évaluation. Les coûts humains entraînés par un mauvais emploi des instruments d'évaluation, la perte d'emploi notamment, peuvent à notre avis être évités par une meilleure utilisation des ressources d'évaluation. Nous tenterons donc de fournir des pistes pour un usage plus adéquat de l'évaluation dans un contexte *d'évaluation sommative.* Cependant, plusieurs de nos recommandations valent aussi pour le contexte *d'évaluation formative.* En fait, nous estimons que les exigences de qualité des évaluations devraient être d'autant plus élevées que leurs conséquences sur le devenir du professeur sont importantes.

Venue de nos voisins américains, l'évaluation par les étudiants est devenue prépondérante dans les prises de décision institutionnelles concernant l'avancement des professeurs. Elle fournit en effet une source et des moyens de mesure facilement accessibles et peu coûteux pour les institutions. Cependant, si elle reçoit la faveur de nombreux administrateurs en raison de sa commodité, elle suscite aussi plusieurs mécontentements de la part des diverses personnes en cause. Les professeurs doutent en effet souvent de la qualité des moyens de mesure utilisés, en l'occurrence, le questionnaire étudiant. D'autre part, les étudiants doutent que l'exercice d'évaluation ait un quelconque impact sur l'amélioration de l'enseignement. Enfin, les administrateurs ne savent pas toujours comment utiliser adéquatement les résultats fournis par les évaluations. Faute de connaissances ou d'information, ces derniers sont appelés à prendre des décisions à partir d'une compréhension partielle des données dont ils disposent.

Nous tenterons donc de cerner ici, à l'aide de la documentation scientifique couverte, les problèmes les plus courants liés à l'évaluation des cours par les étudiants. Nous insisterons surtout sur les problèmes d'*usage* liés au questionnaire étudiant, lesquels, croyons-nous, dépassent en importance ceux qui sont liés aux questions de la *validité.* Nous croyons en effet pouvoir dire que, si les années 80 ont fait l'objet de nombreuses études de

validité, les années 90 ouvrent certainement la voie aux études portant sur l'usage (et le mauvais usage) des questionnaires étudiants par les preneurs de décision. Nous verrons que la question de l'usage renvoie à la notion de *système*, aux différentes personnes qui œuvrent à l'intérieur de ce système, de même qu'aux *responsabilités* qui incombent aux personnes. Après avoir abordé ces questions, nous discuterons de l'emploi des étudiants et des questionnaires étudiants comme *source* et *instrument* d'évaluation. Il sera alors aussi question des contextes formatif (amélioration de l'enseignement) et sommatif (promotion du personnel) de l'évaluation. Cette première partie de l'article conclura en faveur d'une vue pluraliste des sources et des moyens d'évaluation. Nous présenterons ensuite une série d'études de validité touchant aux effets simples et aux effets d'interaction liés à l'évaluation par les étudiants. Finalement, nous émettrons quelques pistes à suivre pour un meilleur usage des évaluations étudiantes par les preneurs de décision.

Le système d'évaluation

Selon Theall et Franklin (1990), l'évaluation étudiante peut être conçue comme un *système* qui inclut à la fois les activités de recherche et les pratiques liées à celles-ci. Une bonne compréhension du système d'évaluation implique, d'une part, une connaissance de la théorie générale de l'évaluation et, d'autre part, une connaissance approfondie des variables en cause, de leur effets et des relations qui existent entre celles-ci.

Par ailleurs, *le système d'évaluation opère dans le contexte d'une politique et d'une philosophie institutionnelles qui met à contribution à la fois les professeurs, les étudiants et les administrateurs.* Ce système peut, en ce sens, affecter la satisfaction et la motivation du professeur et, ultimement, le rendement de l'étudiant. Les chances de succès du processus d'évaluation dépendent, entre autres choses, de la qualité des moyens utilisés pour recueillir l'information recherchée et de la capacité des personnes concernées à se servir des résultats d'évaluation de façon appropriée. Malheureusement, il arrive assez souvent que les résultats des évaluations soient divulgués sans que des précautions nécessaires aient été prises pour s'assurer de la qualité de l'information produite et de la qualification des personnes qui l'utilisent.

Les usages inappropriés peuvent survenir dans plusieurs situations et pour plusieurs raisons: lorsque les politiques d'utilisation des questionnaires sont mal définies, lors de la procédure de collecte de données,

lorsque les résultats sont disséminés sans information préalable sur leurs usages ultérieurs, lorsque leurs destinataires éventuels sont peu ou mal préparés, etc. Ils dépassent donc la question de la vérification de la fidélité et de la validité des instruments de mesure (Theall et Franklin, 1990).

L'usage et la responsabilité

Selon Theall et Franklin (1990), la plupart des recensions d'écrits effectuées dans le domaine de l'évaluation de l'enseignement par les étudiants (par exemple, celles de Centra, 1989, et de Marsh, 1987) s'entendent sur certains points. Les évaluations par les étudiants sont, dans l'ensemble, multidimensionnelles, sont fiables et stables, dépendent du professeur plutôt que du cours, sont relativement valides et peu affectées par des biais potentiels et sont considérées utiles par les administrateurs, par les étudiants et (à un moindre degré peut-être) par les professeurs.

En partie pour ces raisons, l'évaluation de l'enseignement par les étudiants est devenue monnaie courante dans la plupart des institutions d'enseignement supérieur. Une enquête effectuée par Seldin (1984) montre qu'au début des années 80, 70% des universités et des collèges américains avaient recours aux évaluations par les étudiants. Apparemment, cette tendance n'a fait que s'accroître depuis.

Comment se fait-il dès lors, s'interrogent Theall et Franklin (1990), que l'évaluation par les étudiants suscite toujours autant d'acrimonie? Miller (1987) explique ceci par certaines caractéristiques propres aux systèmes d'évaluation. L'auteur note que ces systèmes évoluent souvent de façon plus ou moins aléatoire et organisée, sont des sources de mécontentement en vertu même de leurs buts, ne peuvent jamais espérer rencontrer une pleine approbation et sont contestables d'un point de vue légal. Les précédents auteurs estiment également que, *lorsque les questionnaires d'évaluation par les étudiants sont employés à des fins de promotion, ceux-ci devraient toujours être utilisés conjointement avec d'autres sources d'évaluation de l'enseignement.*

Toujours selon Theall et Franklin (1990), la recherche dans le domaine de l'évaluation ne devrait pas se limiter uniquement aux questions de mesure mais aussi à celles liées à la pratique et aux usages. La faiblesse des systèmes d'évaluation ne résiderait donc pas tant dans les questions relatives à leur fidélité ou à leur validité, mais plutôt sur le plan de leur utilisation. Aussi, bien que de nombreux chercheurs se soient penchés sur les biais potentiels liés aux évaluations, peu d'entre eux mentionnent les

moyens de faire face à ces biais. *L'étendue des mauvais usages est en fait relativement mal connue, mais ceux-ci ont plus de chances de se propager en l'absence d'une planification et de procédés clairement établis à partir des données scientifiques disponibles.*

Theall et Franklin (1990) mettent en doute le fait que les principaux utilisateurs des résultats d'évaluation, à savoir les administrateurs, les professeurs et les étudiants, aient les compétences nécessaires pour interpréter raisonnablement les résultats des évaluations. Les auteurs remarquent à ce sujet que plusieurs administrateurs prennent, de façon routinière, des décisions concernant leur personnel sans connaître les principes de base de l'évaluation. Leur étude démontre aussi que des pratiques inappropriées surviennent régulièrement.

Franklin et Theall (1990) ont par ailleurs reçu, dans leurs universités respectives, plusieurs demandes d'information émanant des professeurs. D'après leurs observations, les professeurs ne recourent généralement pas, faute parfois de connaissances adéquates, aux instructions et aux renseignements qui sont parfois fournis avec leurs évaluations. Leur étude démontre aussi que *même les mises en garde clairement soulignées dans les instructions ont peu d'effets sur la tendance des usagers à utiliser des données inadéquates ou invalides.* Ils ont aussi trouvé qu'environ *la moitié des personnes interrogées qui avaient eu un rôle à jouer dans les prises de décision concernant le personnel professoral avait des connaissances insuffisantes* des statistiques élémentaires liées à l'évaluation et des biais potentiels pouvant affecter les résultats. Ces données sont également appuyées par les résultats de Harris (1990). Les auteurs insistent donc sur l'importance de concevoir un meilleur système pour rapporter les résultats des évaluations aux personnes concernées.

De plus, la plupart des personnes ressources interrogées dans leur étude acceptent l'idée que les utilisateurs doivent travailler avec des données «moins que parfaites mais suffisamment bonnes». Cette idée recoupe une pensée déjà exprimée par Doyle (1989) et partagée par certains spécialistes en évaluation. Selon Franklin et Theall (1990), cette idée empêche sans doute certains abus d'interprétation de la part des personnes expertes impliquées dans les prises de décisions. Cependant, les nuances que cela suppose ne sont, pour le moment du moins, pas à la portée de tous les preneurs de décision actuels. En conséquence, *la combinaison de données moins que parfaites avec des utilisateurs moins que parfaits peut mener rapidement à des usages inacceptables.* Pour ces raisons, il est urgent que des balises soient mises en place afin que les utilisateurs sachent comment

reconnaître les problèmes de validité et de fidélité, comprennent les limites inhérentes aux données d'évaluation et connaissent des procédures valides d'utilisation des données dans les contextes d'évaluation sommative et formative.

L'évaluation par les étudiants

Les étudiants constituent, à l'heure actuelle, l'une des sources les plus employées pour évaluer la qualité de l'enseignement du professeur. C'est également la méthode d'évaluation la plus controversée (Dilts, Haber et Bialik, 1994). Braskamp et Ory (1994) attribuent à différents facteurs l'accroissement du recours à cette source. Le regain provient, entre autres, d'un mouvement national américain visant l'amélioration de l'enseignement et des demandes croissantes du public pour obtenir des critères objectifs de rentabilité institutionnelle. De plus, les problèmes financiers rencontrés par les institutions demandent des coupures de postes qui doivent se justifier par le plus grand nombre de preuves tangibles possible. Enfin, les évaluations des étudiants sont fréquemment utilisées, non pas parce qu'elles constituent en tant que tel un meilleur type d'évaluation, mais plutôt pour des raisons de commodité. Les évaluations peuvent être recueillies aisément et donnent des résultats chiffrés qui peuvent être utilisés éventuellement dans des procédures légales (Eble, 1984).

Ce type d'évaluation constitue donc pour l'université une occasion de se fixer des objectifs et des critères d'efficacité. Toutefois, la conception nécessairement limitée qu'entraîne toute évaluation au moyen d'un questionnaire (ce dernier contenant un ensemble fermé d'items ou d'énoncés, une vingtaine en général, pour tenter de circonscrire toute la complexité du concept d'enseignement) risque de mener à un point de vue partiel ou partial de l'enseignement (Braskamp et Ory, 1994). Les auteurs mettent aussi en garde les utilisateurs contre l'emploi détourné de cet instrument à des fins de «sondages d'opinion» auprès des étudiants. Selon eux, le questionnaire d'évaluation par les étudiants devrait remplir deux fonctions principales. Une fonction formative, à savoir, celle de fournir des données au professeur afin d'améliorer certains aspects de son enseignement, et une fonction sommative, à savoir, celle de fournir des données aux administrateurs pour leurs prises de décision touchant le personnel et les programmes. Ils admettent cependant qu'*il y a peu de moyens présentement pour coordonner les résultats des évaluations avec les prises de décision administratives.* À ce problème s'ajoute aussi celui déjà discuté de la compétence des responsables du système d'évaluation.

Selon Nadeau (1977) les étudiants sont les premiers agents engagés dans le processus d'enseignement et peuvent, pour cette raison, participer «dans une certaine mesure» à son évaluation. En tant que «bénéficiaires», ils peuvent constituer une source appropriée pour juger particulièrement des aspects suivants (Braskamp et Ory, 1994; Dilts, Haber et Bialik, 1994):

la relation étudiant-professeur;
le comportement professionnel ou éthique du professeur;
la charge de travail de l'étudiant;
les connaissances acquises dans le cours;
l'équité dans l'attribution des notes;
l'habileté du professeur à communiquer clairement et de façon organisée;
le degré d'attention et d'intérêt suscité dans le cours.

Par contre, les étudiants ne sont généralement pas bons juges lorsqu'il s'agit d'évaluer un contenu de cours ou le degré d'érudition du professeur dans son domaine. Les étudiants de premier cycle universitaire ne peuvent juger adéquatement la mise à jour des connaissances du professeur. De même, un étudiant inscrit en première année du baccalauréat ne peut répondre véritablement à une question touchant à la capacité du professeur à intégrer les concepts d'un cours avec ceux du programme dans lequel est inscrit cet étudiant.

En général, les professeurs voient avec suspicion l'évaluation par les étudiants. Ils estiment souvent que celle-ci est biaisée par différents facteurs, comme la difficulté du cours, la note attendue, la motivation de l'étudiant, la taille du groupe, etc. Nous verrons plus loin, dans la section portant sur la validité, que ces craintes sont en partie justifiées. Le scepticisme à l'égard des évaluations provient aussi de leur mauvais usage. En fait, les dangers de mauvaises utilisations seraient encore plus dommageables que l'utilisation d'un mauvais instrument d'évaluation (Franklin et Theall, 1989). Seldin (1993) souligne que les abus d'usage sont fréquents. Parmi les mauvais usages, Centra (1993) mentionne les altérations portées aux questionnaires, les décisions administratives importantes comme la promotion, la permanence ou les augmentations salariales (p. 48), l'établissement de rangs entre les professeurs, etc.

C'est en partie pour ces raisons que Braskamp et Ory (1994) préfèrent l'emploi d'une mesure d'évaluation de la performance des étudiants comme source première d'évaluation du professeur. La description de l'apprentissage de l'étudiant, y compris celle des difficultés rencontrées,

peut, à leur avis, supplanter avantageusement les jugements d'efficacité d'enseignement habituellement utilisés. Une plus grande centration sur les apprentissages de l'étudiant devrait aussi favoriser le dialogue, plutôt que la compétition, entre les membres d'une même faculté. *La perspective des étudiants en ce qui concerne leur propre apprentissage peut, selon ces auteurs, être considérée comme une source hautement crédible et utile dans l'évaluation du professeur.* En somme, la crédibilité des étudiants comme source d'information dépend du type de renseignements demandé et des usages tirés de ces renseignements.

Amélioration de l'enseignement

Selon Centra (1993), l'évaluation par les étudiants peut jouer un rôle dans l'amélioration de l'enseignement à condition de satisfaire aux quatre critères suivants. Le professeur doit *acquérir une nouvelle information* à partir des évaluations; il doit *accorder de l'importance à cette information*; il doit *savoir comment faire pour s'améliorer*; il doit *être motivé à faire des améliorations*.

D'après les études, il semble que les professeurs ont une connaissance plus ou moins juste de la perception des étudiants à leur égard. Ainsi, les corrélations entre les évaluations par les étudiants et l'auto-évaluation du professeur varient, selon les études, entre 0,20 (Centra, 1973) et 0,49 (Marsh, 1982). Les évaluations par les étudiants apportent donc certains renseignements qui peuvent être utiles pour le professeur, surtout lorsqu'il s'agit d'un cours donné pour la première fois. Différentes études rapportent en effet que l'évaluation par les étudiants peut apporter certaines améliorations dans le cas où le professeur reçoit effectivement une nouvelle information quant à son enseignement (Cohen, 1980; L'Hommedieu, Menges et Brinko, 1988). Les études types démontrent que les groupes de professeurs évalués au milieu du trimestre améliorent légèrement leurs évaluations à la fin du trimestre (environ un tiers d'un écart type soit, 13 points percentiles). À l'inverse, les groupes de professeurs qui ne reçoivent pas de «feedback» au milieu du trimestre ne voient pas d'amélioration en fin de trimestre. Par ailleurs, le professeur qui accorde peu d'importance aux évaluations et doute de leur validité sera peu enclin à changer son comportement (Centra, 1993).

Le professeur doit aussi avoir les moyens de changer des aspects de son enseignement. Ainsi, les items des questionnaires doivent être assez spécifiques pour apporter une information qui peut par la suite être convertie en comportements concrets. L'étude de Murray (1985) montre, en ce sens, que les professeurs évalués par des items spécifiques (par

exemple, «le professeur écrit le plan du cours sur le tableau») s'améliorent mieux que les professeurs évalués par des items plus généraux que l'auteur appelle *high-inference items* (par exemple, «les cours étaient bien planifiés»). Par ailleurs, selon la recension d'études de Cohen (1980), les professeurs les plus susceptibles de s'améliorer sont aussi ceux qui ont reçus de l'aide sous forme de consultation et qui démontrent une motivation interne (Centra, 1993).

Enfin, les facteurs motivationnels jouent différemment selon qu'ils sont de nature externe ou interne. Les motivations externes, comme l'accès à la permanence, entraînent des améliorations dans l'enseignement chez plusieurs professeurs non agrégés, mais auraient peu d'influence chez les professeurs déjà agrégés (Murray, 1984). Une étude de McKeachie (1979) démontre cependant que des évaluations négatives répétées ont pour effet de décourager le professeur et de le rendre anxieux. Dans ce dernier cas, les évaluations sont peu susceptibles d'apporter un quelconque bénéfice.

La promotion

Selon une étude effectuée par Marlin (1987) dans deux universités américaines, plusieurs étudiants doutent que leurs évaluations aient un quelconque effet sur les prises de décisions concernant la permanence. Plusieurs croient aussi que le professeur porte peu d'intérêt à l'égard des évaluations. La même étude indique cependant que 70% des directeurs de département interrogés se sont effectivement servis des évaluations par les étudiants dans leurs décisions concernant la permanence ou la promotion des professeurs. Une autre étude effectuée sur neuf campus de l'université de Californie démontre que 78% des professeurs rapportent des changements dans leur enseignement à la suite des évaluations par les étudiants (Outcalt, 1980). La perception des étudiants quant à leur impact reflète donc peu la réalité.

La valeur accordée à l'enseignement et aux évaluations par les étudiants dans les décisions institutionnelles varie d'une université à l'autre. Selon Centra (1977), la plupart des directeurs de département estiment que les évaluations par les étudiants doivent être utilisées pour l'évaluation sommative de l'enseignement. De plus, Murray (1984) montre que les professeurs d'une université canadienne qui ont obtenu leur permanence avaient obtenu au préalable des scores d'évaluation significativement plus élevés que ceux qui ne l'avaient pas obtenue. Toutefois, les évaluations par les étudiants ont eu peu d'effet quant à la promotion des professeurs au rang de titulaire de cette université. Il serait, à ce propos, intéressant de refaire cette vérification dans les universités québécoises et canadiennes d'aujourd'hui.

Murray (1984) a aussi trouvé, dans la même université canadienne, une distorsion entre l'importance relative accordée aux évaluations de l'enseignement par les étudiants (par rapport à d'autres sphères d'activités du professeur) et l'impact de celles-ci sur les décisions administratives. Bien que la productivité en recherche ait été considérée comme valant cinq fois le poids des évaluations par les étudiants, ce dernier critère s'est avéré néanmoins responsable de 12% de la variance dans les décisions de permanence. Ceci signifie que les institutions ou les départements qui donnent «consciemment» plus de poids à l'évaluation par les étudiants augmentent considérablement l'impact de cette source de mesure dans les prises de décision.

Le questionnaire étudiant

L'évaluation par les étudiants repose généralement sur l'administration de questionnaires distribués vers la fin des cours. Selon Marlin (1990), il y aurait plusieurs milliers de ces questionnaires, ce qui rend à peu près impossible toute généralisation quant à leur construction. Par ailleurs, les départements décident souvent de construire leur propre questionnaire sans pour autant avoir une supervision adéquate par des experts en évaluation. Le processus de construction des questionnaires et la qualité de l'évaluation qui en découle varient donc considérablement d'un endroit à l'autre (Dilts, Haber et Bialik, 1994).

En dépit de sa variété, le questionnaire étudiant est un instrument qui se présente généralement sous la forme d'une échelle de cotation (*rating scale*) comprenant un nombre limité et défini de questions ou de dimensions avec un choix de réponses. Les questions correspondent à des dimensions jugées pertinentes pour l'évaluation du professeur par les auteurs des questionnaires. Les dimensions pour évaluer l'enseignement étant multiples, les questionnaires employés sont généralement de nature multidimensionnelle. Par ailleurs, le choix de réponses est habituellement contenu sur une échelle en trois à sept points représentant un continuum allant de: «tout à fait en accord» ou «très important» à «tout à fait en désaccord» ou «pas du tout important». Les différents points de l'échelle correspondent à des valeurs numériques qui sont par la suite compilées dans le but de quantifier les attitudes, les jugements et les perceptions des répondants.

Les questionnaires de type multidimensionnel contiennent le plus souvent des *items spécifiques* qui reflètent un ensemble de construits ou de

dimensions distincts servant à décrire l'*efficacité de l'enseignement.* Les items spécifiques concernent généralement, entre autres, les dimensions suivantes: l'habileté («Le professeur maîtrise bien sa matière»); le rapport («Le professeur est amical»); la structure («Le professeur gère bien son temps en classe»); la difficulté («Le professeur donne des lectures difficiles»); l'interaction («Le professeur facilite les discussions en classe»); le «feedback» («Le professeur tient les étudiants informés quant à leurs progrès»). Les questionnaires contiennent aussi quelques *items généraux* reflétant l'évaluation générale par l'étudiant (par exemple: «Comment jugez-vous l'habileté générale du professeur?»).

La liste suivante présente les catégories d'items que l'on rencontre le plus souvent dans les questionnaires d'évaluation des professeurs. Originalement établie par Feldman (1976, 1984), elle a été quelque peu modifiée par la suite par différents chercheurs (Abrami et d'Appolonia, 1990; Centra, 1993). Les catégories d'items présentées (suivies d'un exemple) ont fait l'objet de nombreuses études de validité.

Tableau 1

Dimensions ayant fait l'objet d'études de validité (classement d'après Feldman, 1976)

1. **Stimulation de l'intérêt**
   Le professeur présente la matière de manière intéressante
2. **Enthousiasme**
   Le professeur démontre du dynamisme et de l'enthousiasme envers sa matière
3. **Connaissance du sujet**
   Le professeur a une connaissance approfondie de sa matière
4. **Préparation et organisation du cours**
   Chaque période de cours était bien préparée
5. **Clarté et compréhension**
   Le professeur résumait ou mettait l'emphase sur les points les plus importants au cours des présentations ou des discussions
6. **Habiletés d'élocution**
   Le professeur s'exprime de façon claire et audible
7. **Niveau de la classe et progrès**
   Le professeur semble s'apercevoir quand les étudiants ne comprennent pas le matériel

Tableau 1, (suite de la page précédente)

8. **Clarté des objectifs de cours**
   Les objectifs de cours du professeur étaient clairs

9. **Pertinence et valeur du matériel de cours**
   Les textes utilisés pour le cours étaient utiles

10. **Pertinence et utilité du matériel supplémentaire**
    En général, j'évaluerais les lectures supplémentaires utilisées comme_____

11. **Charge de travail**
    Les étudiants devaient travailler fort dans ce cours

12. **Résultats perçus**
    Le cours a amélioré mes connaissances générales

13. **Équité dans l'évaluation**
    Les examens reflétaient les aspects importants du cours

14. **Organisation de la classe**
    Les étudiants pouvaient s'exprimer sur la manière de faire les choses

15. **Caractéristiques personnelles**
    Le professeur avait le sens de l'humour

16. **«Feedback»**
    Le professeur faisait des commentaires utiles à propos des travaux et des examens.

17. **Encouragement à la discussion et à la diversité d'opinions**
    Dans cette classe, nous avons essayé de comprendre des points de vue différents

18. **Défi intellectuel et encouragement à la pensée indépendante**
    Ce cours fut pour moi un défi sur le plan intellectuel

19. **Préoccupation et respect pour les étudiants**
    Le professeur était amical

20. **Disponibilité et aide**
    Le professeur était facilement disponible pour consultation, par rendez-vous ou autrement.

21. **Le cours en général** (Centra, 1993; Abrami et d'Appolonia, 1990)
    En général, ce fut un cours efficace

22. **Le professeur en général** (Abrami et d'Appolonia, 1990)
    En général, le professeur était efficace

23. **Divers** (Abrami et d'Appolonia, 1990)
    La dimension, du groupe était satisfaisante

Certains auteurs, dont Dilts, Haber et Bialik (1994) et Marsh (1987), pensent que les différents items constituant les questionnaires devraient recevoir des poids différentiels. Les premiers auteurs remettent en question aussi la pertinence d'utiliser des items généraux (par exemple: «en général, le professeur était efficace») quand la majorité des autres items sont spécifiques. Les poids auraient l'avantage de refléter l'importance relative des items spécifiques par rapport à la qualité générale de l'enseignement. D'après Marsh (1987), les poids pourraient être établis par les professeurs eux-mêmes ou en fonction des résultats des recherches mettant en corrélation les différents items avec l'apprentissage de l'étudiant.

Par ailleurs, il est mentionné à plusieurs endroits, dans les ressources consultées, que l'emploi des items spécifiques pose des problèmes lorsqu'ils sont employés à des fins sommatives. En effet, les contextes d'enseignement sont trop variés pour qu'un ensemble restreint de questions spécifiques puisse en rendre compte de manière exhaustive. Pour cette raison, plusieurs auteurs, dont Abrami et d'Appolonia (1990), estiment que les items spécifiques devraient être éliminés des questionnaires au profit de quelques items généraux seulement. En fait, comme il est démontré dans la section suivante, l'évaluation devrait tenir compte des contextes spécifiques dans lesquels l'enseignement est actualisé. De plus, ces auteurs nient, à l'encontre d'une croyance largement répandue dans les milieux pédagogiques, qu'il y ait des qualités intrinsèques pour décrire un bon enseignement. Selon eux, *un enseignement prend toujours forme dans un contexte et des circonstances particulières qui en affectent la qualité.* Pour cette raison, l'enseignement ne pourrait être défini par un ensemble d'items spécifiques, si nombreux soient-ils.

Problèmes liés à la multidimensionnalité

Abrami et d'Apollonia (1990) soulèvent plusieurs difficultés liées à l'emploi de mesures multidimensionnelles pour évaluer l'enseignement. De telles mesures présupposent à tort, selon eux, que les caractéristiques liées à l'efficacité de l'enseignement sont, dans leur essence, invariables. En effet, plusieurs chercheurs tiennent pour acquis que les qualités importantes d'un bon enseignement sont les mêmes indépendamment des cours, des départements ou des universités considérés. D'autres chercheurs se sont intéressés aussi à trouver les qualités «essentielles» pour évaluer tout type d'enseignement indépendamment de son contexte spécifique. Selon Abrami et d'Appolonia (1990), une telle démarche ignore le fait qu'«un enseignement efficace peut en fait être dépendant d'une situation spécifique» *(situation-specific).*

Les analyses factorielles appliquées aux questionnaires multidimensionnels tiennent aussi pour acquis l'invariabilité des qualités liées à l'enseignement. L'analyse factorielle concerne en effet l'analyse des réponses aux énoncés du questionnaire «à travers» les professeurs, les cours, les départements, etc. Elle permet ainsi d'établir certaines corrélations entre des sous-ensembles d'items. Toutefois, le fait que l'on puisse reproduire les résultats de l'analyse factorielle à travers différentes situations d'enseignement (différents professeurs, différents cours, différents départements) n'implique pas que l'on puisse reproduire les mêmes résultats lorsque l'on emploie différents questionnaires qui prétendent mesurer la même chose. Autrement dit, *même si l'application répétée d'un même questionnaire donne toujours les mêmes résultats pour un même cours, un même professeur ou un même département, ceci ne veut aucunement dire qu'il en va de même lorsque l'on emploie un autre questionnaire.*

S'il était possible de reproduire les résultats des différents questionnaires d'évaluation existants, on devrait s'attendre à ce que les mêmes qualités d'enseignement émergent de chaque questionnaire. Il n'y aurait en fait pas de variabilité entre les différents questionnaires sur le plan des facteurs d'efficacité d'enseignement que ceux-là tentent de mesurer. De plus, si l'on admet que les facteurs d'efficacité d'enseignement sont invariables, il ne devrait pas y avoir de différence dans le type et la proportion d'items mesurant ce construit dans les différents questionnaires. Par exemple, les facteurs contenus dans le questionnaire A devraient être les mêmes que les facteurs contenus dans le questionnaire B ou équivalent à ceux-ci. Comme cela semble être difficilement le cas à partir de l'examen des questionnaires existants, Abrami et d'Appolonia (1990) émettent des doutes importants quant à l'idée de l'invariabilité.

Pour vérifier leur doute, Abrami et d'Appolonia (1990) ont répertorié un ensemble de 43 études ayant déjà fait l'objet d'un examen de validité dans différentes recherches. Le critère de validité retenu dans les différentes études recensées équivaut à la corrélation entre les cotes moyennes des jugements des étudiants et la moyenne de rendement des étudiants à des examens. Selon ce point de vue, les jugements des étudiants sont considérés valides lorsqu'ils reflètent correctement l'impact du professeur sur l'apprentissage de l'étudiant tel que mesuré par le rendement à l'examen. Les auteurs ont donc établi, dans un premier temps, un ensemble de dimensions et d'items fréquemment rencontrés dans différents questionnaires. Ils se sont entre autres basés sur l'étude de Feldman (1976) qui avait répertorié 21 dimensions présentes dans la majorité des questionnaires ayant déjà fait l'objet d'études de validité. Ils ont par la suite ajouté trois dimensions aux

21 dimensions dégagées par Feldman (voir le tableau 1 plus haut). Leur travail consistait ensuite à vérifier dans quelle mesure les différents items contenus dans les différentes dimensions constituant les divers questionnaires étaient ou non invariables.

Les résultats démontrent que toutes les dimensions sont représentées dans tous les questionnaires d'évaluation soumis aux étudiants. Cependant, la fréquence d'apparition des dimensions varie d'un questionnaire à l'autre. Les dimensions qui apparaissent le plus souvent dans tous les questionnaires recensés sont la «clarté» et la «compréhension» de même que le «cours en général». Par contre, l'«enthousiasme du professeur» et le «matériel supplémentaire» apparaissent plus rarement dans les questionnaires.

Par ailleurs, Abrami et d'Appolonia (1990) ont trouvé un faible indice d'uniformité dans les différents questionnaires, ce qui indique que les coefficients de validité sont en fait multidimensionnels ou appartiennent à différentes dimensions. (Un haut indice d'uniformité indique que les coefficients de validité représentent une seule dimension à travers tous les questionnaires). De plus, les écarts types de ces coefficients de validité sont élevés (entre 0,12 et 0,45) ce qui indique qu'il y a peu de consistance dans les dimensions à travers les questionnaires. En fait, seules les dimensions générales (le «cours en général», le «professeur en général») ont des indices d'uniformité raisonnablement élevés. À titre d'exemple pour illustrer ceci, les auteurs mentionnent que parmi les 43 coefficients de validité représentant les corrélations entre les cotes de jugement des étudiants sur la dimension «stimulation de l'intérêt» et le rendement de l'étudiant, seulement sept de ces coefficients relevaient de cette dimension unique. Les 36 autres coefficients de validité étaient basés sur d'autres dimensions de l'efficacité de l'enseignement.

Les résultats d'Abrami et d'Appolonia (1990) confirment donc leur doute à savoir que les questionnaires multidimensionnels d'évaluation de l'enseignement sont en fait variés, c'est-à-dire non uniformes les uns par rapport aux autres. Les items et les dimensions varient d'un questionnaire à l'autre, sauf en ce qui concerne les dimensions générales. Les auteurs fournissent un exemple pour décrire les conséquences possibles de la variance entre les questionnaires.

Supposons que le professeur A utilise le questionnaire A tandis que le professeur B utilise le questionnaire B pour évaluer chacun son enseignement. Supposons aussi que les deux questionnaires contiennent un item

relatif à l' «organisation du cours». Cependant, le questionnaire A place cet item dans la dimension x alors que le questionnaire B place cet item à deux endroits différents, soit dans les dimensions y et z. Dans ce cas, on ne peut pas considérer que les deux questionnaires sont équivalents. L'item en question étant doublement représenté dans le questionnaire B, le professeur évalué par ce questionnaire se trouve doublement favorisé ou défavorisé, selon le cas, par rapport au professeur A évalué seulement une fois sur cet item au moyen du questionnaire A. Étant donné cette différence dans les questionnaires, on ne peut inférer que les différences de résultats obtenus éventuellement chez les deux professeurs soient dues à des différences dans leur enseignement. Elles sont plutôt dues aux différences dans les questionnaires employés.

Les auteurs recommandent donc, pour ces raisons, que seuls les items généraux soient considérés dans les prises de décisions. Ils rejoignent en cela, comme nous le mentionnions plus haut, les suggestions d'autres chercheurs qui déconseillent l'utilisation d'items spécifiques dans les questionnaires d'évaluation lorsque ceux-ci servent à des fins administratives. Les items spécifiques, en raison même de leur particularité, s'appliquent mal à différents cours ou types d'enseignement. Au contraire, les items généraux s'appliquent à la majorité des situations d'enseignement. De plus, Abrami et d'Appolonia (1990) découragent l'emploi de questionnaires multidimensionnels à des fins sommatives. Les critiques que les professeurs adressent à l'endroit de l'emploi de certains items spécifiques dans les questionnaires sont, à la lumière des résultats précédents, fondées et devraient en conséquence être considérées avec tout le sérieux qu'elles méritent.

Pour contourner partiellement ces problèmes, certains auteurs suggèrent d'employer des questionnaires à double volet, l'un contenant des items spécifiques à des fins formatives, et l'autre contenant des items globaux à des fins sommatives (Doyle, 1983). Les items spécifiques devraient être élaborés par le professeur et les résultats devraient lui être retournés sans autres intermédiaires. Les items globaux devraient, quant à eux, être choisis par le département en concertation avec les professeurs. Les résultats pourraient ensuite être retournés au professeur et à l'administrateur du département. Bernard et Trahan (1989) ainsi que plusieurs autres chercheurs suggèrent l'emploi de questionnaires formatifs en cours de session et l'emploi de questionnaires sommatifs à la fin de la session.

Autres problèmes

Le questionnaire étudiant a pour but d'évaluer le processus d'enseignement du professeur. Cependant, on lui reproche souvent, avec raison, d'évaluer davantage le professeur ou sa personnalité que l'enseignement. De plus, des vices de forme ont été soulignés par les auteurs. Ces problèmes seront abordés plus loin en détail dans la section portant sur la validité.

Plusieurs reprochent aussi au questionnaire étudiant d'avoir un contenu restrictif qui donne une vue traditionnelle ou simpliste de l'enseignement (p. ex. : «Quelle note accorderiez-vous à votre professeur?»). De même, d'après Shulman (1993) et Wilson (1988), *les items contenus dans les questionnaires d'évaluation par les étudiants renforcent un point de vue conventionnel de l'enseignement, à savoir celui d'un étudiant passif et d'un professeur actif.* En effet, la plupart des items contenus dans les questionnaires sont centrés sur la performance du professeur plutôt que sur celle de l'étudiant (Centra, 1993). Par exemple, plusieurs items concernent l'habileté du professeur à communiquer («Est-ce que le professeur présentait clairement la matière?»). Cette optique entre en contradiction avec les nouveaux courants pédagogiques qui s'orientent davantage maintenant vers des enseignements de type interactif. D'après Braskamp et Ory (1994), les méthodes de mesure actuelles ont souvent pour effet de décourager l'exploration de nouvelles voies d'enseignement.

Pour ces raisons, entre autres, la majorité des auteurs s'entendent pour dire qu'on ne peut espérer rendre compte de toutes les particularités des différents cours existants par un seul et même questionnaire. En conséquence, il est suggéré d'employer différents types de questionnaires pour évaluer différents types d'activités d'enseignement (Bernard et Trahan, 1989; Braskamp et Ory, 1994; etc.).

Pour la diversité

Les auteurs insistent sur le fait que les sources et les moyens doivent être diversifiés et complémentaires si l'on veut atteindre un plus grand niveau d'équité dans l'évaluation du professeur. Aussi, à l'heure actuelle, la recherche sur la crédibilité ne devrait plus se limiter aux sources et aux moyens les plus courants à savoir, l'évaluation par les étudiants à l'aide de questionnaires multidimensionnels. Il importe, pour l'heure, d'utiliser d'autres sources et d'autres moyens tout aussi valables. En fait, il est préférable de concevoir l'évaluation comme un système dans lequel différents instruments peuvent être utilisés de manière non exclusive. Par exemple, les évaluations par les étudiants combinées avec les appréciations des pairs et

le jugement d'experts dans le domaine du professeur fournissent une évaluation plus riche et plus nuancée qui tient mieux compte des multiples facettes de l'enseignement. De même, des questionnaires d'analyse du matériel pédagogique et des rapports d'auto-appréciation du professeur peuvent contribuer à part entière, à côté du questionnaire étudiant, pour donner une meilleure image du professeur. En fait, plusieurs renseignements sont accessibles aux utilisateurs et méritent toute leur attention.

La validité

La *validité* est un concept relatif plus difficile à cerner que celui de la fidélité. Braskamp et Ory (1994) étendent sa définition pour inclure, en plus des questions habituelles touchant aux instruments ou aux méthodes de collecte de données, celles, plus importantes à leurs yeux, des généralisations et des inférences dégagées à partir des données. Ainsi, *un instrument peut être valide pour un usage donné sans l'être pour un autre usage.* Selon un exemple tiré des mêmes auteurs, les évaluations de l'enseignement par les étudiants peuvent être valides lorsqu'il s'agit de mesurer le degré de clarté de présentation de la matière, sans pour autant être valides lorsqu'il s'agit de mesurer un autre aspect de l'enseignement («la connaissance, par le professeur, du sujet présenté», par exemple).

Une enquête faite dans une université américaine démontre que la plupart des professeurs émettent des doutes sur la validité des évaluations par les étudiants. En effet, plusieurs considèrent que les évaluations sont biaisées par différents facteurs (Franklin et Theall, 1989; Marsh, 1987). L'étude de Marsh établit les pourcentages suivants de professeurs qui s'interrogent sur différents biais possibles dans les évaluations: la difficulté du cours (72%); la note attendue («*grading leniency*») (68%); la popularité du professeur (63%); l'intérêt préalable de l'étudiant pour la matière (62%); la charge de travail (60%); la taille de la classe (60%); les raisons pour prendre le cours (55%); la note à l'examen d'entrée (*GPA*) (53%). Nous verrons plus loin que les intuitions des professeurs sont partiellement fondées d'un point de vue scientifique.

Qu'est-ce qu'un *biais* dans le contexte de l'évaluation? Un biais est une circonstance (situation, événement, facteur) qui affecte indûment les évaluations des professeurs bien que cette circonstance n'ait aucun rapport avec l'efficacité de l'enseignement. Un professeur qui donne intentionnellement de bonnes notes à ses étudiants et qui reçoit en retour de bonnes évaluations par ceux-ci est un exemple de circonstance biaisée. Le

comportement du professeur dans ce cas n'a pas de lien avec l'efficacité perçue de son enseignement. Un autre exemple de situation biaisée serait le cas d'un professeur qui, en raison seulement du fait qu'il donne un cours à un petit groupe d'étudiants, obtiendrait de bonnes évaluations. On conçoit ici facilement le désavantage pour un professeur qui donne un cours à un grand groupe. En raison justement des possibilités de biais inhérentes aux évaluations, les administrateurs doivent être parfaitement au courant des facteurs spécifiques qui affectent les évaluations.

Par ailleurs, il n'est pas suffisant d'observer seulement des facteurs ou des biais isolés. Indépendamment, certaines caractéristiques propres aux étudiants, au cours ou au professeur peuvent ne pas avoir d'effets. Cependant la *combinaison* de plusieurs caractéristiques peut avoir des effets favorables ou défavorables sur les évaluations. Étant donné l'existence de biais connus, des mécanismes doivent être mis en place pour découvrir et contrôler ces effets indésirables. Des méthodes statistiques relativement accessibles et faciles d'emploi sont déjà disponibles.

Plusieurs études ont été menées dans le but d'établir la validité des jugements des étudiants au regard de l'enseignement (Cashin, 1988). Les recherches portant sur l'efficacité ou la compétence d'enseignement ne donnent pas toujours des résultats cohérents. Par ailleurs, la plupart des études sont de type corrélationnel et ne permettent donc pas d'établir avec certitude de rapport de causalité entre les facteurs. Cependant, comme le souligne Centra (1993), l'usage de mesures multivariées devrait permettre de clarifier certains rapports de cause à effet. Nous tenterons de tracer ici seulement quelques résultats généraux relatifs aux effets simples et aux effets d'interaction liés aux évaluations par les étudiants.

Les effets simples

Braskamp et Ory (1994, p.177 à 179) présentent sous forme de tableau l'ensemble des biais potentiels qui ont fait l'objet de plus d'une cinquantaine de recherches depuis les vingt dernières années. Les biais potentiels sont placés sous cinq catégories: l'administration du questionnaire, la nature du cours, le professeur, les étudiants et l'instrumentation. Des effets positifs simples, c'est-à-dire des biais qui amènent une meilleure évaluation, ont été démontrés dans les cas suivants.

Tableau 2

Ensemble des biais potentiels et des biais positifs ayant fait l'objet d'études de validité [a]

Administration
- lorsque les évaluations sont signées par les étudiants * [b]
- lorsque les évaluations sont effectuées en présence du professeur *
- lorsque les évaluations sont faites après que les étudiants aient été informés de leur rôle dans la promotion du professeur *
- lorsque les évaluations sont faites durant une classe régulière plutôt qu'en période d'examens *

Nature du cours
- lorsqu'il s'agit d'un cours au choix plutôt que d'un cours obligatoire *
- lorsqu'il s'agit d'un cours avancé ou de deuxième et troisième cycles *
- lorsque le nombre d'étudiants dans la classe est peu élevé (effet minimal *)
- les meilleures évaluations sont données en ordre décroissant aux disciplines suivantes: arts, sciences humaines, sciences sociales et biologie, administration, informatique, mathématiques, génie, sciences physiques

Professeur
- les professeurs reçoivent des évaluations plus élevées que les chargés de cours
- lorsque le professeur est chaleureux et enthousiaste *
- lorsque le professeur fait de la recherche (effet minimal)
- le sexe du professeur n'a pas d'effet
- l'âge, le rang de professeur et le nombre d'années d'expérience ont peu d'effet

Étudiants
- lorsque les étudiants s'attendent à recevoir des notes élevées *
- lorsque les étudiants ont un intérêt préalable par rapport au cours *
- lorsque les étudiants sont inscrits dans une «majeure» plutôt que dans une «mineure»*
- lorsque l'étudiant est du même sexe que le professeur (masculin-masculin ou féminin-féminin) (effet minimal *)
- la personnalité de l'étudiant n'a pas d'effet

Instrumentation
- lorsque seulement les deux points extrêmes sur l'échelle de dimension sont précisés *
- l'utilisation d'une échelle en six points donne des résultats plus fiables qu'une échelle en cinq points.
- le nombre d'items formulés négativement n'a pas d'effet
- l'emplacement des items sur le formulaire n'a pas d'effet

a. D'après Braskamp et Ory, 1994.
b. L'astérisque indique un biais positif.

L'ensemble des recherches portant sur les effets simples de certains facteurs sur les évaluations par les étudiants démontre les tendances générales suivantes. Certaines caractéristiques du professeur, des étudiants et des cours affectent les évaluations: l'expressivité du professeur, l'indulgence dans l'attribution des notes, la méthode d'enseignement, la taille du groupe, l'intérêt préalable pour la matière et la discipline académique. La question à savoir si les orientations pédagogiques peuvent être considérées comme un biais, c'est-à-dire un facteur qui affecte indûment les évaluations par les étudiants, est présentement débattue. Ce facteur peut en effet facilement se confondre avec le facteur «personnalité», lequel est considéré par plusieurs comme un biais.

En raison de ces biais, Centra (1993) recommande que *les décisions touchant au personnel ne se prennent seulement qu'après un examen minutieux de tous les facteurs interférant avec l'évaluation du professeur.* La mise en circulation de résultats semblables à ceux qui sont présentés ici constitue, à notre avis, un premier pas pour faciliter des prises de décision plus éclairées. Ces résultats constituent en effet un instrument précieux pour la découverte des biais relatifs aux évaluations des professeurs. Selon nous, chaque unité départementale devrait de plus voir à faire ses propres études de validité afin de dégager les biais qui peuvent exister dans leur cas particulier. Centra recommande aussi que *les décisions administratives se prennent seulement après avoir considéré plusieurs cours différents donnés pendant plusieurs années.* De plus, les différentes corrélations obtenues dans les études de validité soulignent l'importance de faire conjointement appel à plusieurs moyens d'évaluation de l'enseignement et de ne pas s'en tenir seulement au questionnaire d'évaluation de l'enseignement par les étudiants.

On peut donc dire que si les étudiants constituent une source d'information dans l'évaluation des professeurs, la prise en considération de certaines précautions est nécessaire. Selon Dilts, Haber et Bialik (1994), on ne peut ignorer l'ensemble des études qui ont démontré l'existence de biais significatifs rattachés aux questionnaires d'évaluation par les étudiants. Des conclusions valides et des décisions justes ne peuvent être faites qu'après avoir effectué des vérifications sur les plans de la construction, de l'administration et de l'interprétation des résultats du questionnaire. *Plusieurs auteurs préviennent donc les usagers contre la tentation d'utiliser le questionnaire étudiant comme la seule mesure valable d'évaluation de l'enseignement.* Les dimensions relatives à l'évaluation d'un bon enseignement sont simplement trop nombreuses et diverses pour être contenues dans un seul instrument (Dilts, Haber et Bialik, 1994).

Les effets d'interaction liés au sexe

Les études portant sur les effets d'interaction posent la question à savoir *si un effet (ou l'absence d'effet) relevé pour une variable donnée dans un contexte ou une situation donnés demeure le même lorsqu'une ou plusieurs variables additionnelles sont prises en considération.* On peut ainsi se demander si la corrélation nulle trouvée entre le sexe du professeur et l'évaluation globale de ce professeur est maintenue, ou au contraire change, à travers différentes situations. Une autre manière de chercher des interactions possibles consiste à voir si certaines dimensions de l'évaluation (ou certains items spécifiques des questionnaires d'évaluation) sont plus fortement corrélées que d'autres à l'évaluation globale du professeur en fonction d'un facteur donné.

La présente recherche a conduit à l'examen de plusieurs études d'interaction. Leur nombre étant potentiellement très grand, nous avons opté pour la présentation de quelques-unes seulement. Notre choix s'est basé, d'une part, sur la quantité de recherches existantes et, d'autre part, sur notre jugement personnel quant à la pertinence des facteurs pris en considération. Il nous a paru intéressant de présenter ici les grandes lignes des résultats et des interprétations sur les effets d'interaction du sexe du professeur avec différents facteurs simples ou combinés. La plupart des études d'interaction recensées prennent pour variable dépendante l'évaluation globale du professeur telle que faite par les étudiants. La variable dépendante correspond aussi parfois à l'apprentissage de l'étudiant (Hull et Hull, 1988). Ces variables sont mises en corrélation avec les facteurs considérés pour découvrir des effets possibles.

L'absence d'effets simples liés aux facteurs «sexe du professeur» et «sexe de l'étudiant» ne saurait signifier que ces variables n'interviennent pas dans les évaluations. C'est plutôt dans l'interaction de ces facteurs entre eux et avec d'autres facteurs que résident des différences. Ainsi, le fait que les évaluations globales des professeures féminines et des professeurs masculins ne diffèrent pas significativement et le fait que les évaluations des étudiants et des étudiantes ne diffèrent pas non plus significativement sont tous deux indépendants du fait que *des dimensions spécifiques peuvent jouer de manière différentielle en fonction du sexe.* Certains facteurs agissent de façon plus importante que d'autres dans les évaluations globales des professeurs en fonction du sexe (Feldman, 1993, p. 178).

L'examen du facteur «orientation pédagogique» de même que celui du facteur de «la personnalité» du professeur (avec lequel il est souvent

confondu) indique des différences en fonction du sexe des professeurs. La métaanalyse de Feldman (1993) montre à ce sujet que les professeures féminines obtiennent généralement de meilleures cotes pour les dimensions typiquement associées aux femmes, par exemple, «la sensibilité et l'attention au progrès des étudiants». Ceci est compatible avec l'idée déjà avancée par Bennet (1982) et reprise par Resnick Sandler (1991), à l'effet que *les étudiants féminins et masculins ont des attentes différentes selon le sexe du professeur.* Les professeures féminines ont plus de chances d'être perçues comme devant donner du «support» aux étudiants. Les étudiants et les étudiantes risquent alors plus de leur demander certains traitements de faveur et de s'attendre à obtenir gain de cause (par exemple, demander une extension de délai dans la remise d'un travail).

Par ailleurs, l'étude de Bennet (1982) indique que si les étudiants sont d'accord de façon informelle pour juger la professeure féminine comme étant «disponible», ce jugement ne transparaît pas dans leur évaluation formelle du professeur. Même si les professeures féminines passent effectivement plus de temps avec les étudiants, comme les rapports de ces derniers sur le nombre de visites au professeur l'indiquent, les professeures féminines ne sont pas pour autant jugées plus disponibles que les professeurs masculins dans les évaluations. Ceci démontre que la clientèle étudiante serait plus exigeante à l'égard des professeures féminines quant à leur disponibilité. De plus, les études de Simeone (1987) montrent que les professeures féminines seraient davantage portées à donner du soutien aux étudiants tout en étant par ailleurs conscientes des abus auxquels ceci peut donner lieu de la part des étudiants.

Plusieurs études indiquent l'existence d'un *double standard* quant aux exigences requises chez les professeures féminines et les professeurs masculins (Kierstead, D'Agostino et Dill, 1988; Resnick Sandler, 1991; Statham, Richardson et Cook, 1991). Un même comportement peut être perçu positivement, négativement, ou encore ne pas avoir d'effet en fonction du sexe du professeur. Ainsi, l'étude de Kierstead, D'Agostino et Dill (1988) montre que les évaluations des professeures féminines sont plus fortement influencées par le fait que celles-ci sont «sociables» et «souriantes», alors que ces attitudes affectent peu les évaluations globales des professeurs masculins.

De même, le fait de présenter la matière de façon «non définie» et «ouverte à la discussion» est interprété chez une professeure féminine comme un manque de compétence et est jugé négativement. Cette dernière remarque est compatible avec la conclusion de Bennet (1982) à l'effet que

les étudiants ont tendance à être plus exigeants envers les professeures féminines quant au niveau de «structuration» et d'organisation du cours. Ce niveau doit, selon lui, être plus élevé chez les professeures féminines afin d'obtenir une parité dans les évaluations globales avec les professeurs masculins.

De plus, les comportements «maternels» attendus par la population étudiante à l'égard de la professeure féminine peuvent entrer en contradiction avec d'autres caractéristiques jugées importantes pour un bon enseignement, comme le «dynamisme» ou le «leadership» (Resnick Sandler, 1991). À l'inverse, la professeure féminine qui affiche des caractéristiques plus typiquement masculines peut aussi être mal jugée du fait qu'elle ne se conforme pas aux attentes des étudiants. Par exemple, des exigences intellectuelles élevées de la part de la professeure féminine à l'égard de ses étudiants peuvent être interprétées comme un désir de contrôle. Feldman (1993), reprenant l'idée de Statham, Richardson et Cook (1991), explique à ce sujet que les professeures féminines sont, contrairement à leurs collègues masculins, soumises à une double contrainte (*double bind*).

Enfin, la diversité des méthodes employées dans les études portant sur le sexe est certainement en partie responsable du peu de consistance dans les résultats obtenus (voir Basow et Silberg, 1987). En effet, certaines études montrent que les professeurs masculins reçoivent de meilleures évaluations que les professeures féminines. D'autres études indiquent que les étudiants masculins ont tendance à moins bien évaluer les professeures féminines que les étudiantes (Basow et Silberg, 1987; Brooks, 1982; Kaschak, 1978; Lombardo et Tocci, 1979). Selon ces auteurs, les étudiantes jugent les professeurs masculins et féminins de manière équivalente sur plusieurs dimensions de l'évaluation. Par contre, les étudiants jugent systématiquement mieux les professeurs masculins que les professeures féminines.

Pour résoudre ces contradictions et déterminer clairement l'influence de certains facteurs, Statham, Richardson et Cook (1991) suggèrent de faire des pairages entre les professeures féminines et les professeurs masculins sur la base de leur rang académique, des divisions de cours et des années d'expérience d'enseignement universitaire. Ils rappellent également que les facteurs liés au sexe du professeur et à celui de l'étudiant sont responsables de seulement 4% de la variance dans les évaluations globales des étudiants. Les facteurs liés aux traits de personnalité sont en fait beaucoup plus prépondérants dans les évaluations (Erdle, Murray et Rushton, 1985).

Pour les raisons évoquées, il paraît plus indiqué, à l'heure actuelle, de pousser davantage les recherches sur les effets d'interaction. Les études d'interaction reflètent mieux la complexité des situations réelles d'enseignement. Un professeur masculin n'est pas seulement un professeur masculin, mais aussi un professeur masculin placé dans une classe x, enseignant un cours y, à des étudiants z, etc.

Pistes à suivre pour un meilleur usage des évaluations

Nous proposons maintenant quelques pistes à suivre pour une utilisation plus adéquate des évaluations. Nous avons vu en effet que les problèmes d'usage sont probablement encore plus importants que les problèmes liés aux mesures. Ces derniers sont, par ailleurs, considérablement bien documentés par rapport aux problèmes liés à l'utilisation des résultats d'évaluation.

Les pistes suggérées ici prennent la forme de points à suivre pour tirer un meilleur profit des évaluations par les étudiants pour un usage général. Nous sommes grandement redevables à Centra (1993) pour avoir établi les grandes lignes directrices.

Déterminer la manière dont les résultats seront utilisés

Les professeurs et les administrateurs doivent connaître préalablement et clairement l'usage prévu pour les évaluations. Les usages peuvent être divers: l'amélioration de l'enseignement, l'octroi de la permanence, la promotion, l'augmentation de salaire. De plus, les personnes concernées doivent savoir qui aura accès aux résultats des évaluations et comment ceux-ci sont reliés aux contrats qu'elles ont avec l'institution. *Les étudiants doivent aussi être informés des usages prévus pour l'évaluation et du sérieux de toute cette entreprise* (Ory, 1990). L'institution qui omet de donner ces renseignements encourt le risque que le professeur prenne lui-même l'initiative de le faire. En prenant la responsabilité d'expliquer aux étudiants les conséquences de leur évaluation, l'institution évite au professeur des situations gênantes (par exemple, celle de dire qu'il sera congédié ou qu'elle sera congédiée de son poste si il ou elle n'obtient pas sa permanence après cinq ans). *Les étudiants doivent en effet savoir qu'ils sont un élément clé important dans le processus entier d'évaluation du professeur.*

Utiliser plusieurs sources d'information

La pluralité des sources d'information comporte le recours à plusieurs ensembles de résultats et à un nombre suffisant d'évaluateurs. Indépendamment de l'usage prévu pour l'évaluation par les étudiants (sommatif ou formatif), celle-ci ne saurait être suffisante en elle-même (Centra, 1993). De nombreux auteurs ont en effet suggéré d'autres sources d'information comme les rapports des collègues ou les auto-évaluations. L'emploi de plusieurs indicateurs pour évaluer l'efficacité de l'enseignement du professeur est surtout important lors des prises de décision.

De plus, les décisions administratives doivent être prises seulement après l'examen des résultats de plusieurs évaluations. Les administrateurs doivent donc considérer la tendance générale des évaluations plutôt que l'effet d'un cours particulier, par exemple. Selon Centra (1993), une telle tendance devrait commencer à se dégager lorsqu'un même cours a été donné au moins cinq fois. Par ailleurs, pour être représentative, *l'évaluation du professeur doit couvrir, si possible, différents types de cours* (par exemple, des cours dispensés en vue du baccalauréat et des cours de deuxième et troisième cycles, des cours optionnels et des cours obligatoires).

Bien qu'il soit généralement admis que l'évaluation doive se faire de façon systématique, c'est-à-dire, pour tous les professeurs et tous les cours donnés, une telle démarche peut entraîner certaines difficultés. Ory (1990) souligne à cet effet que l'évaluation systématique peut, en certains cas, refréner le désir chez certains professeurs d'expérimenter de nouvelles voies d'enseignement. De même, les études démontrent que les évaluations par les étudiants sont affectées par certaines caractéristiques du cours (Centra, 1993). *Les administrateurs doivent donc examiner ces caractéristiques et les mettre en relation avec les résultats des évaluations.* La taille du groupe, la méthode d'enseignement, le sujet du cours, sont des facteurs qui ont peu d'effet lorsque pris isolément mais qui peuvent avoir un impact considérable lorsque combinés (jusqu'à 30 points percentiles, selon Centra). Aussi, bien que la considération d'un ensemble de cours soit en général préférable à celle de quelques cours particuliers, le professeur devrait néanmoins avoir le droit d'éliminer certaines de ses évaluations lorsque les circonstances le justifient. Ory (1990) recommande à ce sujet que les administrateurs respectent les demandes des professeurs.

Le nombre absolu d'étudiants et la proportion d'étudiants qui répondent effectivement au questionnaire sont tous les deux importants dans l'établissement de la fidélité des évaluations. Celle-ci s'accroît avec le nombre d'étudiants-répondants. Pour cette raison, les évaluations basées

sur moins de dix étudiants ou moins de cinq classes doivent être jugées avec précaution. En effet, l'impact d'un ou de deux étudiants mécontents est davantage dilué dans un grand groupe.

Utiliser les données comparatives avec précaution

Les données comparatives permettent de situer le professeur par rapport à une moyenne d'évaluations. Selon Centra (1993), les comparaisons entre professeurs doivent être basées sur un grand échantillon de professeurs et, de préférence, sur plusieurs années. Certains questionnaires commerciaux, comme le *SIR* et l'*IDEA* aux États-Unis, publient à cet effet des données comparatives basées sur les résultats d'ensemble des institutions qui utilisent ces questionnaires. Parfois aussi, l'université utilise ses propres données locales pour établir des comparaisons. Les administrateurs doivent, dans ce cas, éviter de faire des comparaisons entre les professeurs d'un département qui compte moins de quinze membres.

Par ailleurs, le fait que les résultats des évaluations soient présentés sous forme chiffrée donne l'impression d'une plus grande précision, ce qui n'est pas le cas en réalité. Il n'y a probablement pas de différence significative entre un professeur qui se situe au quinzième percentile et un autre qui se situe au seizième percentile. De plus, *l'établissement de rangs percentiles signifie «par définition» que la moitié des professeurs se situe sous le cinquantième percentile.* Centra (1993) propose donc que les professeurs soient plutôt classés en quatre ou cinq groupes, et seulement après que plusieurs évaluations aient été faites. Enfin, l'auteur souligne que l'établissement de rangs entre les professeurs d'un même département peut entraîner une atmosphère de compétition contreproductive.

Utiliser les items généraux pour les prises de décision

Plusieurs auteurs (Abrami et d'Appolonia, 1990; Braskamp et Ory, 1994; Centra, 1993) recommandent aux institutions d'utiliser les items généraux contenus dans les questionnaires d'évaluation plutôt que les items spécifiques qui sont moins fiables. Les résultats aux items généraux ont une corrélation plus élevée avec d'autres mesures de la qualité d'enseignement, comme la performance des étudiants aux examens. Cette recommandation est supportée par de nombreuses recherches (Abrami, 1985, 1988, 1989; Cashin et Downey, 1992). Selon Abrami (1989), *les questionnaires devraient en fait contenir seulement des items généraux lorsque ceux-ci sont utilisés à des fins sommatives.* Des exemples d'items généraux utiles à l'évaluation sont des questions comme: «Comment jugeriez-vous les capacités générales du

professeur?»; «comment jugeriez-vous la qualité de ce cours?», «combien avez-vous appris dans ce cours par rapport à d'autres cours?».

Les items spécifiques ne sont pas recommandés à des fins sommatives parce qu'ils représentent de façon insuffisante les multiples aspects de l'enseignement (voir la discussion plus haut sur la «multidimensionnalité»). De plus, ils sont peu valides pour évaluer une variété de cours, de professeurs, d'étudiants et de contextes différents. Abrami (1989) conteste l'emploi d'items spécifiques, par exemple, ceux touchant à la qualité de l'interaction professeur-étudiant dans tous les contextes d'enseignement: petit ou grand groupe, cours axés sur des discussions, ateliers pratiques, cours magistral, etc. Quelle est la pertinence, en effet, de la question à savoir si «le professeur était amical envers les étudiants» dans les classes à haut taux d'inscription? On a également de la peine à imaginer ce que serait un cours où une centaine d'étudiants seraient «encouragés à discuter». *Les situations d'enseignement sont trop diversifiées pour prétendre que toutes les caractéristiques d'enseignement sont également importantes.* Pourtant, c'est précisément l'idée sous-entendue par tous les questionnaires d'évaluation (Abrami et d'Appolonia, 1990).

Pour assurer la validité des questionnaires, leurs concepteurs doivent s'assurer que ces instruments recueillent un ensemble suffisant de jugements d'étudiants. Ces jugements doivent de plus être en rapport direct avec l'ensemble des comportements effectifs du professeur dans sa classe. Les questionnaires ne doivent, par conséquent, contenir ni trop d'items associés avec un comportement donné ni trop peu d'items associés avec un autre type de comportement donné. De plus, les questionnaires doivent contenir des items pertinents et également appropriés à chacune des situations possibles d'enseignement.

Plusieurs auteurs (Abrami, 1989; Abrami et d'Appolonia, 1990; Cohen, 1981) mettent en doute l'efficacité des questionnaires utilisant des items spécifiques (questionnaires multidimensionnels) en raison du fait qu'ils sont peu reliés à l'apprentissage des étudiants. La métaanalyse effectuée par le dernier auteur démontre des corrélations faibles ou nulles entre les dimensions représentées dans les questionnaires: «rapport», «interaction», «feedback», «évaluation», «difficulté» et «apprentissage des étudiants». Par contre, les corrélations augmentent pour les dimensions générales (p. ex., «cours en général», «professeur en général». Par conséquent, *la validité de certains items spécifiques ne paraît pas supportée par des résultats de recherche.* De plus, la validité des items spécifiques est susceptible de varier en fonction des différentes situations d'enseignement. Ainsi, le fait de

savoir qu'un biais potentiel n'a pas d'effet sur un item spécifique offre peu d'assurance à l'effet que ce même biais n'aura pas d'effet sur un autre item spécifique (Abrami, d'Appolonia et Cohen, 1990).

L'utilisation des items spécifiques devrait en fait être réservée au contexte formatif de l'évaluation. Ils sont en effet plus utiles que les items généraux pour le professeur qui veut améliorer certains aspects de son enseignement (Braskamp et Ory, 1994; Centra, 1993). Le professeur peut aussi, ou devrait à tout le moins, ajouter ses propres questions sur le questionnaire d'évaluation standard et encourager les commentaires écrits ou verbaux de la part des étudiants. Cependant, il doit garder à l'esprit qu'*il n'est pas possible ni même désirable de satisfaire à toutes les demandes des étudiants* (Centra, 1993).

Pour les raisons évoquées, Abrami et d'Appolonia (1990) croient qu'*il est peu réaliste de s'attendre à ce que les administrateurs arrivent à des décisions pleinement fondées concernant la qualité de l'enseignement à partir des questionnaires existants.*

Utiliser une procédure standard pour administrer le questionnaire

Selon Dilts, Haber et Bialik (1994), il existe aujourd'hui des controverses quant au nombre minimum d'étudiants constituant un groupe représentatif et quant au meilleur moment pour faire les évaluations. De même, la question du nombre de fois qu'un professeur doit être évalué reste plus ou moins bien définie. D'après ces auteurs, il serait préférable d'évaluer le cours du professeur seulement une fois par année lorsque les évaluations servent à des fins administratives. Le fait d'évaluer tous les cours à tous les trimestres peut créer des effets secondaires indésirables. Les étudiants ont tendance à répondre de façon moins sérieuse lorsqu'ils remplissent de façon répétée les mêmes questionnaires et ne voient pas d'amélioration subséquente dans l'enseignement du professeur (Centra, 1993; Marlin, 1990). De plus, certaines caractéristiques de l'enseignement du professeur peuvent être absentes du questionnaire. Le changement cyclique de questionnaires peut, en ce sens, apporter un élément de solution. Il permet d'avoir un plus grand éventail d'items et permet de mieux représenter le professeur.

Il est nécessaire d'employer une procédure d'administration standard lorsque l'opération d'évaluation sert à des prises de décision administratives. Les formulaires doivent être remplis par les étudiants en classe, avant la période d'examen final et en l'absence du professeur. La procédure habituelle requiert en effet que le professeur quitte la classe et

qu'un étudiant fasse la surveillance. Il existe aussi d'autres façons de faire; le protocole du test PERPE de Gagné (1977) permet au professeur de rester dans la classe, tout en suivant certaines conditions particulières. De même, la surveillance peut être assurée par un conseiller pédagogique. Les étudiants doivent rester anonymes et ne pas prendre plus de quinze minutes pour répondre. Le professeur peut, s'il quitte la classe, rappeler que l'évaluation doit se faire de façon individuelle. Les discussions entre les étudiants pendant le processus d'évaluation constituent une faute de procédure qui peut mener à l'invalidation des résultats. Ory (1990) souligne à cet effet qu'il peut être difficile, voire impossible, de savoir si la procédure a été correctement observée, à moins que le professeur en soit informé par un étudiant ou qu'une personne neutre ait dirigé les opérations. La personne désignée pour l'évaluation doit indiquer le nombre de questionnaires remplis et non remplis. Elle doit signer et cacheter l'enveloppe contenant tous les questionnaires en présence de la classe. L'enveloppe doit être remise au bureau assigné en présence de plusieurs témoins. Le professeur peut s'assurer personnellement que l'enveloppe a été correctement cachetée. Cette façon de faire évite les mauvaises manipulations de la part des différentes personnes en cause. Enfin, les résultats présentés sous forme quantifiée ainsi que les commentaires écrits des étudiants doivent être retournés au professeur, selon des délais raisonnablement courts, à la fin du cours.

Des dispositions particulières doivent être également prises s'il advient que l'une ou l'autre des parties (étudiants, professeurs, ou administrateurs) soupçonne une mauvaise manipulation des résultats des évaluations. L'institution doit, en ce cas, prendre les accusations avec tout le sérieux qu'elles méritent. Pour cela, elle doit avoir prévu des règlements quant aux procédures pour porter plainte, pour déterminer un mauvais agissement et, éventuellement pour imposer des pénalités.

Donner une chance au professeur de répondre aux évaluations

Le professeur doit avoir l'opportunité d'expliquer les circonstances dans lesquelles le cours s'est déroulé. Certains questionnaires peuvent ne pas refléter adéquatement les méthodes utilisées dans un cours donné (Centra, 1993). Par exemple, les questionnaires traditionnels peuvent ne pas bien représenter une méthode d'enseignement non traditionnelle. Le professeur peut, si bon lui semble, inclure avec ses évaluations une description écrite des circonstances liées à un ou quelques cours particuliers dont les résultats semblent aberrants.

Conclusion

Le but de cet article était d'identifier quelques-unes des failles inhérentes au système d'évaluation tel qu'il est encore couramment utilisé dans les universités et de recommander quelques moyens pour y remédier. Nous avons vu que certaines directives méthodologiques peuvent être suivies afin d'assurer un meilleur usage des évaluations. Il est aussi ressorti clairement que plusieurs sources et moyens d'information doivent être mis à contribution de façon à fournir un portrait plus fidèle de l'enseignement du professeur. Les évaluations par les étudiants au moyen de questionnaires doivent donc être conçues seulement comme un élément d'information dans l'évaluation générale de la compétence d'enseignement. Ceci est particulièrement important dans un contexte sommatif. Par ailleurs, nous avons souligné que les items généraux fournissent une meilleure base de comparaison entre les professeurs sur le plan d'un département ou d'une faculté. Enfin, nous ne saurions aussi trop insister sur le fait que les différentes personnes en cause dans l'évaluation, professeurs, administrateurs et étudiants, devraient connaître clairement les buts et les usages des résultats de l'évaluation. Ceci devrait avoir pour effet d'augmenter leur intérêt et leur sentiment de responsabilité à l'égard de ce processus.

L'évaluation des professeurs par les étudiants constitue une source d'information valable pour juger d'une dimension de l'efficacité de l'enseignement: la perception des étudiants quant à la performance du professeur. Elle est, en ce sens, une des composantes du système complexe que constitue l'évaluation de la qualité de l'enseignement. Toutefois, comme toute autre donnée, nous avons vu que les évaluations des étudiants peuvent être rassemblées de façon plus ou moins appropriée. Aussi, les institutions qui dépendent largement de celles-ci pour leur prise de décisions doivent établir des mécanismes qui les mettent à l'abri de leur mauvais usage.

Les conséquences de plus en plus lourdes qu'on a actuellement tendance à dégager des évaluations des professeurs par les étudiants amènent de nouveaux problèmes d'équité. Ceux-ci concernent, entre autres, la détresse que peut occasionner la perte d'un emploi. Il paraît donc urgent que les institutions prennent des mesures adéquates et des dispositions claires pour s'assurer d'un usage adéquat des évaluations (Bernard, 1992; Ory, 1990). Ceci peut se faire, dans un premier temps, par une meilleure mise en circulation des données de la recherche qui traitent du

domaine de l'évaluation par des personnes désignées responsables du fonctionnement du système.

RÉFÉRENCES

Abrami, P. C. (1985). Dimensions of effective college instruction. *Review of Higher Education,* 8, 211-228.

Abrami, P. C. (1988). SEEQ and Ye shall find: a review of Marsh's «Students' evaluation of university teaching». *Instructional Evaluation,* 9, 19-27.

Abrami, P. C. (1989). How should we use student ratings to evaluate teaching. *Research in Higher Education,* 30(2), 221-227.

Abrami, P. C. et d'Apollonia, S. (1990). The dimensionality of ratings and their use in personnel decisions. *In M.* Theall et J. Franklin (Éds), *Student ratings of instruction: issues for improving practice*. new directions for teaching and learning, n° 43. (p. 97-111). San Francisco: Jossey-Bass Publishers.

Abrami, P. C., d'Apollonia, S. et Cohen, P.A. (1990). The validity of student ratings of instruction: what we know and what we don't. *Journal of Educational Psychology,* 82, 219-231.

Basow, S. A. et Silberg, N, T. (1987). Student evaluations of college professors: Are female and male professors rated differently? *Journal of Educational Psychology,* 78(3), 308-314.

Bennet, S.K. (1982). Student perceptions and expectations for male and female instructors: evidence relating to the question of gender bias in teaching evaluation. *Journal of Educational Psychology,* 74(2),170-179.

Bernard, H. et Trahan, M. (1989). Les principales composantes d'une politique d'évaluation de l'enseignement. *Mesure et Évaluation en Éducation,* 11(4), 61-84.

Bernard, H. (1992). *Processus d'évaluation de l'enseignement supérieur. Théorie et pratique*. Laval (Québec): Éditions Études Vivantes.

Braskamp, L. A. et Ory, J. C. (1994). *Assessing faculty work: enhancing individual and institutional performance*. San Francisco: Jossey-Bass Publishers.

Brooks, V. (1982). Sex differences in student dominance behavior in female and male classrooms. *Sex Roles,* 8(7), 683-690.

Cashin, W. E. (1988). *Student ratings of teaching: a summary of the research.* Manhattan: Center for faculty evaluation and development, Kansas State University.

Cashin, W. E. et Downey, R. G. (1992). Using global student rating items for summative evaluation. *Journal of Educational Psychology,* 84(4), 563-572.

Centra, J. A. (1973). Effectiveness of student feedback in modifying college instruction. *Journal of Educational Psychology,* 65(3), 395-401.

Centra, J. A. (1977). *How universities evaluate faculty performance: a survey of departments heads*, GREB Research report N° 75-5bR. Princeton, N.J.: Educational Testing Service.

Centra, J. A. (1989). Faculty evaluation and faculty development in higher education. *In* J.C. Smart (Éd.), *Higher Education Handbook of Theory and Research (*vol. 5.) New York: Agathon Press.

Centra, J. A. (1993). *Reflective faculty evaluation: enhancing teaching and determining faculty effectiveness*. San Francisco: Jossey-Bass Publishers.

Cohen, P. A. (1980). Effectiveness of student rating feedback for improving college instruction: a meta-analysis of multisection validity studies. *Research in Higher Education,* 13(4), 321-341.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: a meta-analysis of multisection validity studies. *Review of Educational Research,* 51(3), 281-309.

Dilts, D.A., Haber, L. J. Bialik, D. (1994). *Assessing what professors do: an introduction to academic performance appraisal in higher education.* (ch. 5: The Evaluation of Teaching in Universities, p. 47-63). Westport, Connecticut: Greenwood Press.

Doyle, K.O. (1983). *Evaluating teaching.* Toronto: Lexington Books, D.C. Health and Company.

Doyle, K. O. (1989). *Report on an academic's visit to pragmatica*. Paper presented at the annual meeting of the American Educational Research Association. San Francisco.

Eble, K.E. (1984). New directions in faculty evaluation. *In* P. Seldin (Éd.), *Changing practices in faculty evaluation: a critical assessment and recommendations for improvement*. San Francisco: Jossey-Bass.

Erdle, S., Murray, H.G. et Rushton, J. P. (1985). Personality, classroom behavior, and college teaching effectiveness: a path analysis. *Journal of Educational Psychology,* 77(4), 394-407.

Feldman, K. A. (1976). Grades and college students' evaluations of their courses and teachers. *Research in Higher Education,* 4(1), 69-111.

Feldman, K.A. (1984). Class size and college students' evaluations of teachers and courses: A closer look. *Research in Higher Education,* 21(1), 45-116.

Feldman, K. A. (1993). College students' views of male and female college teachers: Part II- Evidence from students' evaluations of their classroom teachers. *Research in Higher Education* , 34(2), 151-191.

Franklin, J. et Theall, M. (1989). *Who reads ratings: knowledge, attitudes, and practices of users of students ratings of instruction*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Franklin, J. et Theall, M. (1990). Communicating student ratings to decision makers: design for good practice. *In M.* Theall et J.Franklin, J. (Éds). *Student*

ratings of instruction: issues for improving practice. New Directions for Teaching and Learning, n° 43. (p. 75-93). San Francisco: Jossey-Bass Publishers.

Gagné, F. (1977). *Questionnaire PERPE supérieur: Manuel technique*. Montréal: Institut national de la recherche scientifique, INRS-Éducation, Presses de l'Université du Québec.

Harris, S. (1990). *Have users utilized instrumentation recommendations*?. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Hull, D.R. et Hull, J. H. (1988). Effects of lecture style on learning and preferences for a teacher. *Sex Roles,* 18(7/8), 489-496.

Kaschak, E. (1978). Sex bias in students' evaluations of college professors. *Psychology of Women Quarterly,* 2(3), 235-243.

Kierstead, D., D'Agostino, P. et Dill, H. (1988). Sex role stereotyping of college professors: bias in students' ratings of instructors. *Journal of Educational Psychology,* 80(3), 342-344.

L'Hommedieu, R., Menges, R. J. et Brinko, K.T. (1988). *The effects of student ratings feedback to college teachers: a meta-analysis and review of research.* Unpublished manuscript, Center for the Teaching Professions, Northwestern University.

Lombardo, J. et Tocci, M. (1979). Attribution of positive and negative characteristics of instructors as a function of attractiveness and sex of instructor and sex of subject. *Perceptual et Motor Skills,* 48(2), 491-494.

Marlin, J. W. (1987). Student perceptions of end of course evaluations. *Journal of Higher Education,* 58(6), 704-716.

Marlin, J.W. (1990). Student evaluations of instruction: decipher the message. *In* P. Saunders et W.B. Walstad, *The Principles of Economics Course. A Handbook for Instructors*, New York: McGraw-Hill Book Company.

Marsh, H. W. (1982). SEEQ: A reliable, valid and useful instrument for collecting students evaluations of university teaching. *British Journal of Educational Psychology,* 52(1), 77-95.

Marsh, H. W. (1987). Students' evaluations of university teaching: research findings, methodological issues, and directions for future research. *International Journal of Educational Research,* 11, 253-388.

McKeachie, W. J. (1979). Students ratings of faculty: a reprise, *Academe,* october, 384-397.

Miller, R. I. (1987). *Evaluating faculty for promotion and tenure.* San Francisco: Jossey-Bass.

Murray, H. G. (1984). The impact of formative and summative evaluation of teaching in north american universities. *Assessment and Evaluation in Higher Education,* 9(2), 117-132.

Murray, H. G. (1985). Classroom teaching behaviors related to college teaching effectiveness. *In* J. G. Donald and A.M. Sullivan (Éds.), *Using research to improve teaching*. New Directions in teaching and Learning, n° 23 (p. 21-34). San Francisco: Jossey-Bass.

Nadeau, G. (1977). *Student evaluation of instruction: the rating questionnaire*, If teaching is important... The Evaluation of Instruction in Higher Education (p. 73-128) Clarke Irwin and Company Limited.

Ory, J. C. (1990). Student ratings of instruction: ethics and practice. *In* Theall et Franklin, J. (Éds). *Student ratings of instruction: issues for improving practice.* New Directions for Teaching and Learning, n° 43 (p. 63-74) San Francisco: Jossey-Bass Publishers.

Outcalt, D.L. (1980). *Report of the task force on teaching evaluation.* Berkeley: University of California Press.

Resnick Sandler, B. (1991). Women faculty at work in the classroom, or, Why it still hurts to be a woman in labor. *Communication Education,* 40(1), 6-14.

Seldin, P. (1984). *Changing practices in faculty evaluation: a critical assessment and recommendations for improvement.* San Francisco: Jossey-Bass.

Seldin, P. (1993). How college evaluate professors: 1983 versus 1993. *AAHE Bulletin*, october, p. 6-8, 12.

Shulman, L.S. (1993). *Displaying teaching to a community of peers.* Presentation at the First American Associate of Higher Education (AAHE) Conference on Faculty Roles and Rewards, San Antonio, Texas, January 30.

Simeone, A. (1987). *Academic women: working towards equity*. South Hadley, Mass.: Bergin and Garvey Publishers.

Statham, A., Richardson, L. et Cook, J. A. (1991). *Gender and university teaching: a negotiated difference*. Albany, N.Y.: State University of New York Press.

Theall, M. et Franklin, J. (1990). Student ratings in the context of complex evaluation systems. *In* Theall, M., Franklin, J. (Éds). *Student ratings of instruction: issues for improving practice. New Directions for Teaching and Learning,* n° 43 (p. 17-34) San Francisco: Jossey-Bass Publishers.

Wilson, T. C. (1988). Student evaluation of teaching: a critical perspective. *Review of Higher Education*, 12(1), 79-95.