

L'instrumentation en évaluation

Marie-Claire Dauvisis

Volume 29, Number 1, 2006

L'actualité de la recherche en évaluation

URI: <https://id.erudit.org/iderudit/1086967ar>

DOI: <https://doi.org/10.7202/1086967ar>

[See table of contents](#)

Publisher(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (print)

2368-2000 (digital)

[Explore this journal](#)

Cite this article

Dauvisis, M.-C. (2006). L'instrumentation en évaluation. *Mesure et évaluation en éducation*, 29(1), 45–66. <https://doi.org/10.7202/1086967ar>

Article abstract

The social interest in evaluation develops an important ingeniery with many new tools to communicate some results, presented often without neither explanation nor precaution. The construction of instruments appears like a good idea to attract evaluators and teachers to the formation. These facts have to be examined by the evaluation research: they procure it new questions and wide its domain and its traditional points of views.

L'instrumentation en évaluation

Marie-Claire Dauvisis

Équipe PATRE – ENFA Toulouse

MOTS CLÉS: Évaluation, recherche, demande sociale, ingénierie, outils, résultats, validité, instrumentation, modèles, références

Pour répondre à la demande d'évaluation et fournir des résultats, souvent communiqués sans nuances, il se développe toute une ingénierie avec une instrumentation importante; par ailleurs, la construction d'outils d'évaluation apparaît comme une excellente entrée pour conduire les évaluateurs et les formateurs vers la formation. Ces faits méritent d'être examinés par la recherche en évaluation et lui fournissent nombre de nouveaux questionnements, de nature à élargir fortement son champ et ses perspectives traditionnelles.

KEY WORDS: Evaluation, research, social needs, ingeniery, tools, results, validity, instrumentation, models, references

The social interest in evaluation develops an important ingeniery with many new tools to communicate some results, presented often without neither explanation nor precaution. The construction of instruments appears like a good idea to attract evaluators and teachers to the formation. These facts have to be examined by the evaluation research: they procure it new questions and wide its domain and its traditional points of views.

PALAVRAS-CHAVE: Avaliação, investigação, exigência social, engenharia, utensílios, resultados, validade, instrumentação, modelos, referências

Para responder à exigência da avaliação e fornecer resultados, muitas vezes comunicados sem explicações nem precauções, desenvolve-se toda uma engenharia com uma extensa instrumentação; além disso, a construção de utensílios de avaliação surge como um excelente meio para conduzir os avaliadores e os formadores para a formação. Estes factos merecem ser examinados pela investigação em avaliação, fornecem-lhe novas questões, alargando muito o seu campo e as suas perspectivas tradicionais.

Actuellement la demande sociale d'évaluation, ou plutôt de résultats attendus des évaluations, s'accroît. La tentation est grande, car parfois assortie de financements importants, d'accepter de telles commandes et de fournir ainsi des résultats sans être très attentifs aux qualités et limites des outils déployés. Face à une telle prolifération, un besoin de clarification s'impose et la recherche se trouve ainsi mobilisée autour de l'instrumentation de l'évaluation : elle peut y trouver nombre d'importantes questions à explorer pour y apporter quelques éclairages.

Modèles et usages d'une pratique sociale

Dans le cadre du colloque annuel de l'ADMEE-Europe à Aix-en-Provence, en 2001, concernant le lien entre évaluation et formation, un atelier de débats regroupait des communications sous le thème « outils et méthodes » ; à l'automne suivant, lors du congrès d'actualité de la recherche en éducation et en formation, au symposium de l'ADMEE-Europe animé par Anne Jorro, une analyse des communications et des enjeux sous-jacents pour la recherche a été proposée (Dauvisis, 2001). Un premier constat s'imposait : les conceptions de l'évaluation sous-jacentes aux divers apports sont multiples et varient selon les enjeux, les usages et les besoins, les ingénieries se multiplient sans toujours préciser leurs références.

Un mot, des pratiques aux références multiples

De même que, dans les pratiques scolaires, les évaluateurs, dont la conscience professionnelle ne peut être mise en doute, acceptent mal de voir contestés les résultats qu'ils produisent, n'admettant pas que leurs systèmes de valeurs ne soient pas universellement partagés par d'autres collègues qui, comme eux, accompliraient « sérieusement » leur tâche d'évaluation des apprentissages, les conceptions de l'évaluation sont moins unanimement partagées qu'une première approche superficielle pourrait le laisser croire. Les définitions de l'évaluation sont multiples et son champ sémantique s'étend sur de multiples registres : il n'est pas toujours aisé de discerner les différents sens attribués dans la réalité des pratiques et dans les discours qui les accompagnent,

Note de l'auteure : L'auteure est responsable de l'équipe PATRE (Unité propre de recherche « professionnalité, activités, travail et représentation des enseignants), équipe d'accueil doctoral de l'École doctorale CLESCO de l'Université Toulouse-le-Mirail, associée au DEA des Sciences de l'éducation « EFI » (éducation, formation et insertion). Toute correspondance peut être adressée par courriel à l'adresse suivante : [marie-claire.dauvisis@educagri.fr].

créant souvent plus de confusion que de clarification, dans la mesure où les conceptions sous-jacentes ne sont pas toujours clairement définies. L'enjeu, l'usage, les habitudes et le contexte sont des déterminants forts pour la plus ou moins grande légitimité et pour l'implicite de telles références.

Bien des chercheurs en évaluation se sont appliqués, depuis les dernières décennies, – et divers dictionnaires en éducation en témoignent (De Landsheere, 1979; Legendre, 1993) – à tenter de différencier l'évaluation selon ses fonctions en adjoignant à ce substantif des qualificatifs particuliers. Il reste cependant que, dans l'usage courant ou dans celui de certaines présentations de recherche ou de pratiques, le terme générique d'évaluation recouvre des conceptions spécifiques, soit liées à une école universitaire donnée, soit simplement fixées par l'acception que chacun s'en donne dans la situation qui l'intéresse, sans toujours s'assurer qu'une telle définition soit unanimement partagée. C'est une source de confusion indéniable; elle n'est pas spécifique à ce domaine: d'autres termes du langage de l'éducation connaissent bien cette difficulté, dans la mesure où ils appartiennent aussi au langage courant. Pour ne prendre qu'un exemple: actuellement, le terme de compétence se répand à peu près dans tous les milieux, tant professionnels qu'éducatifs, et la recherche a bien du mal à en fixer un usage acceptable qui le conceptualise tant soit peu et le rende opératoire.

Pour définir l'évaluation, des distinctions tendent à s'effectuer dans les milieux de la recherche: les uns séparent définitivement contrôle d'évaluation (Ardoino & Berger, 1986), d'autres souhaiteraient relier l'évaluation à la recherche du sens (Bonniol & Vial, 1997; Vial, 2001), d'autres réservent à l'évaluation formative l'usage du mot évaluation (par exemple, Weiss, 1991 et 1994), d'autres ne parlent d'évaluation que dans une approche éducatrice (départements de mesure et d'évaluation des universités canadiennes), d'autres enfin en termes économétriques d'efficacité de comparaison d'effets (IREDU). Ces divers sens se retrouvent dans les pratiques et les institutions du terrain éducatif; quant à l'approche commune de l'évaluation des apprentissages, elle reste fortement connotée aux épreuves, à la certification et à la notation, c'est-à-dire à la production d'un résultat communicable et socialement accepté. Ce modèle de simple production des résultats n'est-il pas finalement le plus attendu dans la demande sociale?

Il s'agit, certes, d'un vieux débat, mais, on le constate, toujours d'actualité bien que latent: de fait, n'appartiendrait-il pas à l'essence elle-même de l'évaluation, cette pratique sociale liée à une activité humaine si spontanée et si naturelle qu'elle n'a plus à être interrogée? L'évaluation, en effet, mobilise les

valeurs des systèmes de références et chacun se rattache à une école précise, et dans la mesure où celle-ci constitue pour lui «le» modèle – ou qu’il n’en connaît pas d’autres! –, il tend à la poser comme universelle, et les confrontations sont souvent esquivées; dans ce domaine aussi, c’est-à-dire au travers de l’usage du concept lui-même, l’évaluation peut se révéler comme un lieu et un enjeu de pouvoir, une recherche de légitimation et d’identification.

Pour définir ce qui détermine cette «opération de lecture orientée de la réalité» (Hadji, 1997, p. 119) et sortir de la confusion des langues dans le Babel qui règne actuellement dans le domaine de l’évaluation, un tel foisonnement pose à la recherche une première question de terminologie¹ ou d’affichage explicite des théories, des paradigmes (De Ketele, 1993) et des valeurs de référence: il s’agit non seulement d’une stabilisation consensuelle des concepts, mais surtout de leur usage, au moins par les chercheurs, à défaut d’en pouvoir espérer une diffusion généralisée dans les usages communs de la pratique.

Dans les Sciences de l’éducation, cette question de conceptualisation des pratiques est un des dilemmes épistémologiques importants: soit une distance est prise, qui confine à la rupture entre les acquis des recherches et les pratiques qui se développent dans un usage social, soit on vise à réduire cette distance, et le risque de confusion permanente entre chercheurs et autres acteurs de l’éducation demeure car les mots ne prennent pas pour tous le même sens; mais dénommer par le même concept des démarches dont les intentions, les projets et les objets respectifs sont, de nature, très peu semblables ne va pas sans poser question.

Usages et enjeux dans des contextes divers

Si l’on accepte la définition de l’évaluation comme processus de recueil d’information en vue d’une décision à prendre (De Ketele & Roegiers, 1993; Stufflebeam, 1980), la nature et les enjeux de la décision influencent le processus lui-même (Lecoite, 1997): certaines intentions à l’initiative de l’action ne sont pas directement repérables, se révèlent contradictoires, ou recèlent des projets non avouables, elles piloteront cependant de manière plus ou moins implicite tout le processus et ses résultats. La recherche ne peut ignorer cet état de fait et impose de situer l’ensemble du processus dans son contexte et dans son projet.

Évaluer des politiques d'éducation invite à une approche sociologique qui permet de «situer la problématique [...] dans le champ des rapports sociaux macro et microscopiques», en précisant notamment les conditions d'une mise en relation avec les décisions politiques, les perspectives internationales, l'offre d'éducation en fonction du développement économique et social de l'environnement, le rôle de l'évaluation institutionnelle dans le processus de gestion de l'école (Solaux, 2000, p. 28).

Évaluer une innovation nécessite toute une réflexion préalable sur les enjeux de celle-ci pour les divers acteurs et pour les institutions où elle se situe, sur la hiérarchie des objectifs visés, sur les bénéfices secondaires latents et sur les pertes éventuelles réelles.

Évaluer un programme (Nadeau, 1984) implique de déterminer les partenaires en cause, les raisons d'introduction de ce programme, les motifs de la commande d'évaluation, les évolutions du contexte environnant, etc., de connaître les moyens réellement disponibles, de repérer les caractéristiques des formés concernés et de leurs formateurs, de préciser les acquisitions visées et les modifications attendues de la réalisation de ce programme, etc., de pouvoir comprendre les enjeux perçus par les principaux acteurs à la fois pour le programme et pour son évaluation (Figari, 1994), etc.

Évaluer une organisation éducative ou un établissement scolaire (Crahay, 1994) impose de prendre en compte les intérêts et les projets des acteurs concernés ainsi que les orientations fixées pour le pilotage du système éducatif.

Évaluer des apprentissages suppose de percevoir comment les évalués se représentent la portée de cette situation, pour eux et pour leur avenir (CRESAS, 1986; Rayou, 2001), le risque qu'ils encourent (Monteil, 1989), l'intérêt qu'ils peuvent y trouver, etc., de repérer les modèles de l'apprentissage sous-jacents ainsi que les processus cognitifs (Depover & Noël, 1999), la hiérarchie et la nature des objectifs réellement en jeu, etc.

L'inventaire pourrait ainsi se poursuivre sans que l'exhaustivité puisse être envisagée. Cependant, sans ce souci permanent de la prise en compte de la spécificité du contexte et des jeux d'acteurs qui se mettent en place autour de l'évaluation, c'est la validité de la démarche qui peut devenir contestable, conduisant à des conclusions parfois entachées de lourds contresens ou dont la portée ne peut être validée par les acteurs les plus touchés.

Bien que cette étape préliminaire se révèle totalement déterminante pour donner sens à l'évaluation et à ses résultats, on constate qu'elle est souvent négligée et que le projet d'évaluation n'est pas toujours explicité avant

l'engagement dans une démarche de collecte d'information pour la production de résultats tant soit peu assurés et leur communication; le choix de méthodes et d'outils pertinents, valides et fiables en dépend cependant. Ainsi, le discrédit ou la méfiance parfois exprimés vis-à-vis de projets ou de conclusions d'évaluation peuvent-ils se nourrir facilement des soupçons liés à ce déficit de clarification du contexte et des enjeux. Comme toute pratique sociale, l'évaluation se situe dans un contexte de légitimation.

Ainsi une autre question de recherche autour de l'instrumentation de l'évaluation est relative à la signification d'une action située (*cf.* Casalfiore & Paquay, 1998), elle vise la compréhension de ses résultats dans le contexte et avec les moyens de leur production, leurs limites de validité, les possibilités et la capacité de généralisation. C'est là une question liée à l'épistémologie spécifique des recherches en sciences humaines, et plus particulièrement en éducation, qui mettent en jeu des interactions en contexte, inscrites dans une temporalité éphémère, dans des conditions jamais reproductibles et impliquant des acteurs doués d'initiatives, bien souvent assez imprévisibles.

L'instrumentation de l'évaluation : une entrée possible en formation

Proposer, en formation d'enseignants, de construire des outils pour l'évaluation en classe (grilles, consignes, exercices, etc.) attire de nombreux candidats pour des sessions qui y sont consacrées et s'avère un moyen relativement efficace (Vial, 2001) pour un public relativement peu demandeur de formation (Dauvisis, 2003). Cet objectif rejoint, en effet, les préoccupations quotidiennes des enseignants et leur «rapport instrumental» à la formation (Cauterman, Demailly et al., 1999); il ne s'y cantonne cependant pas car, au travers de la construction de situations et d'instruments d'évaluation par les questionnements divers qu'elle renvoie sur la formation, cet objectif conduit alors progressivement les enseignants concernés à adopter «un rapport professionnel à la formation» (*ibid.*): il s'agit alors d'une démarche pilotée par la dynamique d'un projet professionnel personnel, inscrit dans la durée et prenant en compte le contexte d'exercice. Ainsi des enseignants fortement imprégnés du modèle du magistère charismatique ou de l'artisan (Altet, 1994) sont conduits à adopter une posture d'enseignant réflexif (Paquay, 1994; Schön, 1983) et à remettre en cause personnellement leurs pratiques pédagogiques usuelles.

Ne retrouverait-on pas là, d'une certaine manière et par rapport aux enseignants, l'usage fréquent et bien connu de l'évaluation dans les pratiques scolaires qui est d'en faire un moyen pédagogique pour inciter les élèves à apprendre : les examens et les différentes épreuves sont un levier puissant de l'organisation des curricula et de leur mise en œuvre (Delory, 1996). L'évaluation constitue souvent une bonne clé de lecture des pratiques de formation (Dominice, 1979); l'analyse de pratiques d'évaluation, l'élaboration d'une instrumentation devient aussi un puissant moyen de formation pour les formateurs : rendre des comptes sur ses propres pratiques conduit à les réinterroger et à se remettre en question. En effet, comme en témoignent, lorsque leurs avis et réactions sont pris en compte, les différents acteurs concernés par des évaluations d'innovations et de programmes, toutes les démarches plus ou moins instrumentées de collecte d'information nécessaire à l'évaluation constituent souvent l'origine d'autoquestionnements et d'autoévaluations (Vial, 1993), propices à une évolution personnelle et à l'émergence de besoins de formation. L'organisation de la communication de résultats, les séances de validation des conclusions par les partenaires en cause peuvent également représenter de réelles occasions de formation : il y a là, en effet, une imbrication forte entre formation et évaluation qui ne saurait être trop soulignée.

Une nouvelle question pour la recherche en évaluation mérite d'être posée autour de l'instrumentation des pratiques éducatives d'évaluation, celle du lien établi, affiché ou caché, entre formation et évaluation et des effets attendus de l'évaluation sur la formation, c'est-à-dire en ce qui a trait à la rétroinformation sur la modification des conceptions et des pratiques des acteurs visés (Charlier, 1998). Cela suppose le développement de recherches autour des représentations professionnelles de la formation et de l'évaluation mises en jeu dans l'instrumentation de l'évaluation, à la fois par les évaluateurs et par les « bénéficiaires » de cette action, voire les commanditaires d'évaluations.

Une attitude responsable : répondre de ses actions et de ses choix

Dans un monde fortement piloté par l'économie, l'évaluation correspond à une demande sociale importante, pour ne pas dire unanime : il faut rendre des comptes sur ses actions et sur l'usage des investissements. Perçue comme bien immatériel, l'éducation n'échappe plus désormais à cette logique économique et, en période de limitation des ressources, des choix d'investissements et de moyens sont impératifs, visant une rentabilité maximale de leurs effets : l'efficacité est devenue l'injonction sociale et une des valeurs

fondamentales du développement. Il devient, dès lors, indispensable de prévoir et de pouvoir rendre des comptes : cela constitue non seulement un souci pour les décideurs et les politiques qui doivent justifier leurs décisions, notamment budgétaires, mais c'est aussi, comme en témoignent nombre de débats autour de l'école, celui de leurs électeurs, ces contribuables, véritables « consommateurs d'école » qui enjoignent à la formation de produire des résultats et « du concret » ! L'éducation, fondée sur la transmission et l'acquisition d'une culture générale, apparaît désormais comme insuffisante lorsqu'elle ne permet pas la promotion, ou seulement l'insertion sociale et professionnelle, lorsqu'elle n'ouvre pas la possibilité de réorientations professionnelles ou qu'elle conduit à des exclusions, génératrices de violences qui mettent en danger la paix sociale et fragilisent ses équilibres.

De telles interrogations traversent désormais toutes les sociétés, des moyens sont recherchés et mis en œuvre pour tenter d'y répondre : le recours à l'évaluation en est une conséquence, dans la mesure où celle-ci est souvent présentée comme la panacée, seule capable de fournir des résultats apparemment fondés et, par là même, socialement acceptables. En France, à l'Éducation nationale, une direction du Ministère lui est actuellement consacrée et un Haut Comité de l'évaluation a été mis en place pour assurer, avec l'Inspection générale, de telles fonctions dans le système éducatif. Une véritable « culture de l'évaluation »² tend ainsi à se développer dans tous les domaines, accompagnée d'une ingénierie foisonnante, liée à un véritable marché de l'évaluation avec son jeu de concurrence, et qui, sous des apparences de rationalité, peut se réduire à accompagner un simple rite social, si ce n'est même à entretenir une illusion mystificatrice, pour un profit commercial délibéré.

La recherche se voit, en conséquence, saisie d'une nouvelle injonction concernant l'analyse et la qualité des méthodes et des outils, développés pour satisfaire les commandes et les projets d'évaluation ; elle est également sollicitée pour contribuer à l'élaboration d'outils, munis des qualités intrinsèques minimales, ainsi que pour l'élaboration de modèles permettant l'analyse des pratiques correspondantes.

L'évaluation : recherche de rationalité et rigueur

La rigueur : une réalité, une croyance, un postulat implicite ?

Fournir des résultats, voilà l'enjeu majeur et attendu de l'évaluation, mais les conditions de leur production sont bien peu souvent remises en question, surtout s'ils paraissent crédibles ou permettent de conforter des décisions plus ou moins déjà connues. Comme le soulignait J. Houssaye lors du colloque ADMEE d'Aix-en-Provence :

On sent une certaine fascination pour l'évaluation en tant que mode d'une régulation juste de l'action : ce qui veut dire que l'évaluation serait dans la rigueur et l'action dans l'aveuglement. [...] L'évaluation serait le moyen de faire vrai pour établir le bien. On aurait donc une espèce de transférabilité du résultat par une rigueur de l'évaluation, ce qui veut dire que l'évaluation serait un moyen pour oublier la démarche de production de l'évaluation. La démarche de production d'une action dans le domaine de la formation pourrait être oubliée au profit d'une transférabilité du résultat, car celui-ci est garanti par une évaluation rigoureuse.

De tels propos ne font que traduire la croyance commune en la rigueur de l'évaluation, fondée sur sa capacité à produire des résultats communicables et non sur les moyens et les conditions qui permettent de les produire. Si une telle croyance fonde bien des pratiques et des conceptions de l'évaluation, elle ne saurait sans examen soigneux être admise par les chercheurs qui remettent en question à la fois dans ses fondements culturels et dans ses conséquences : c'est aujourd'hui un des objets centraux de la recherche en évaluation.

Les outils de collecte d'information et leurs résultats

Un foisonnement d'outils et de méthodes

Un rapide recensement de ce qui était présenté dans les diverses communications regroupées dans l'atelier consacré à ce thème lors du colloque d'Aix-en-Provence, permet de relever de multiples outils et méthodes. On y trouvait :

- des constructions d'outils d'évaluation et de dispositifs d'évaluation formative en formation des enseignants ;
- des modalités et contenus de formation à l'évaluation ;
- des outils technologiques, dans leur usage de supports d'échanges entre enseignants ;
- des démarches d'évaluation de projets ou d'innovations ;
- des dispositifs de recherche et des techniques de traitement des données ;
- des résultats de recherches évaluatives, des expérimentations ;

- des indicateurs de réussite des formations ;
- des pratiques d'évaluation/formation : rédaction de mémoire, rapport de stage, soutenance, entretien de recrutement, élaboration de dossier, etc. ;
- des épreuves externes ou construites dans le cadre d'un projet, des référentiels ou grilles de critères pour l'autoévaluation ;
- des mises en œuvre d'outils formalisés : portfolios, plans d'étude pour l'instrumentation de l'évaluation formative ;
- des produits technologiques d'ingénierie concrétisant une démarche d'évaluation : simulateur, outil de validation d'acquis, etc.

Une telle liste manifeste à l'évidence une absence de rigueur dans la définition même de ce qu'est un outil ou une méthode d'évaluation et conduit au constat que, tant dans les milieux de la recherche que dans ceux de la pratique, tout ou presque semble accessible à une telle qualification. De fait, malgré le titre de l'atelier, comme en témoigne cet inventaire, les présentations et analyses d'outils d'évaluation eux-mêmes étaient l'exception ; par ailleurs, dans la plupart des présentations de recherche ou de pratique, les outils de recueil d'information n'étaient pas même donnés en annexe des communications et, en conséquence, aucun débat ne pouvait s'engager sur les valeurs implicites qui les justifiaient, ni sur leurs qualités, pas plus, évidemment, que sur les biais éventuels qu'ils étaient susceptibles d'introduire.

Un constat s'impose : la pratique de l'évaluation semble oublier le rôle fondamental de ces divers aspects dans les résultats produits ; les recherches et les publications de vulgarisation en ces domaines ont pourtant, depuis longtemps, sensibilisé à cette difficulté (Abernot, 1988 ; De Landsheere, 1974 ; Noizet & Caverni, 1979 ; Piéron, 1963). Faut-il conclure à une véritable inculture en matière d'évaluation ? Il apparaît plutôt que la croyance est bien ancrée aussi bien en la rigueur qu'en l'équité (Dauvisis, Millot, Frossard & Droyer, 2003) « naturelles » de l'évaluation, et en conséquence les pratiques d'évaluation n'ont nul besoin de prouver la qualité de leurs outils, dans la mesure où elles produisent des résultats communicables et que, socialement, on leur attribue la valeur d'une vérité.

La recherche établit ce constat : elle ne peut s'y résigner. Dans la mesure où elle a produit et communiqué des résultats établis depuis plusieurs décennies, la nouvelle question qui lui est posée est celle de la diffusion et de la prise en compte des acquis de la recherche dans les milieux de la pratique et des moyens de formation adaptés pour déstabiliser des croyances préjudiciables à

la rigueur de la pensée et pour remettre en question l'activité évaluative elle-même (Figari & Achouche, 2001). Toutefois, si l'évaluation ne relève pas de ce type de rigueur mais acquiert sa valeur de vérité dans un autre registre qui est celui de la légitimité culturelle des pratiques sociales (Dauvisis et al., 2003), c'est un autre champ de recherche, celui de la « traduction » des résultats de recherche, qui doit se développer pour la recherche en évaluation, comme cela a pu être fait dans le domaine des sciences (*cf.* Latour, 1992, 1995) et, en Sciences de l'éducation, avec la recherche sur les transferts d'innovation dans le système éducatif (Cros, 2000).

Qu'est-ce qu'un outil? quelle fonctionnalité?

Un outil, selon la définition du dictionnaire Larousse, c'est un « objet fabriqué dont on se sert pour exécuter un travail manuel ou mécanique ». Certes, avec l'évaluation, on ne se situe ni dans un travail manuel, ni dans un travail mécanique; mais cette définition permet de souligner que c'est le travail à accomplir qui justifie la fabrication de l'outil et finalise son élaboration, et de l'étendre, comme on le fait déjà, en utilisant le concept d'outils d'évaluation. L'utilisation et l'utilité déterminent les choix d'outils et leur élaboration, elles sont souvent surdéterminées par les conditions d'usage et les habitudes et normes en vigueur: la faisabilité de l'évaluation limite parfois les prétentions de perfection des outils; néanmoins, sans une clarification préalable des objectifs visés, il devient problématique, voire très aléatoire, de définir un outil adapté.

Par ailleurs, aucun outil ne peut être considéré comme totalement neutre: sa construction et son organisation révèlent les conceptions sous-jacentes de ses auteurs. Si construire des outils se révèle un support efficace de formation, en analyser ou en déconstruire pourrait jouer un rôle semblable; ce n'est cependant que très rarement envisagé. On le sait, depuis les premières études de Bloom et de ses collaborateurs dans la fin des années 1950 (Bloom, 1969), par exemple, dans le domaine des questions d'évaluation d'examen, les taxonomies ont largement montré que toutes les questions sur un même contenu n'étaient pas équivalentes: selon la façon dont la question était posée, les réponses impliquent des processus cognitifs de complexité peu comparable; de fait, ce ne sont pas les mêmes objectifs qui sont évalués. Ces références semblent parfois oubliées par les divers concepteurs ou utilisateurs d'outils, qui, construisant ou utilisant un outil particulier, n'en précise pas toujours la fonctionnalité, pour leurs intentions et dans les limites du contexte de mise en œuvre.

L'instrumentation de l'évaluation favorise l'usage d'outils divers et variés. Ainsi, identifier la fonctionnalité de l'outil et révéler les valeurs dont il est porteur apparaissent conjointement comme une autre question posée aux chercheurs en évaluation, par le foisonnement actuel de l'ingénierie avec les diverses techniques et outillages plus ou moins sophistiqués qui s'y multiplient.

Des outils pour quelles généralisations ?

L'élaboration ou le choix de l'outil et son adaptation à la situation apparaissent donc comme une étape fondamentale de toute procédure d'évaluation. Diverses solutions peuvent se présenter :

- soit des batteries d'outils – parfois plus ou moins validés ou standardisés – sont disponibles pour des situations semblables ou voisines et, dans ce cas, il est indispensable de construire des critères de choix pour sélectionner les outils les mieux adaptés à la situation spécifique (*cf.* Peretti, 1998);
- soit, à partir d'une claire définition de la fonctionnalité attendue dans le cadre de l'usage particulier, il reste à élaborer un outil possédant les qualités requises;
- soit, enfin, des habitudes ont fixé des usages qui ne sont plus remis en question et les outils correspondants peuvent être utilisés, sans que soit toujours vérifiée leur validité pour les évaluations envisagées.

Si l'on souhaite ne pas négliger un minimum de rigueur scientifique, il faut cependant reconnaître que les données à recueillir pour une évaluation ne se satisfont pas de n'importe quel moyen de collecte ou de traitement et que les outils disponibles ou usuels ne se révèlent pas d'emblée adaptés : c'est ce que montrent clairement désormais certaines études évaluatives, qui commencent à remettre en cause des modèles de traitement des données un peu trop systématiquement utilisés dans l'approche économétrique (Jarousse & Piquée, 2001). Dans la recherche, l'analyse critique des outils et de leur fonctionnalité conduit alors à l'élaboration de nouveaux outils mieux adaptés (Bressoux, Leroy-Audouin & Coustère, 1998). La démarche de référentialisation (Figari, 1994) correspond à une préoccupation semblable : pour donner sens à l'évaluation d'une action, le référentiel n'est pas donné *a priori*, mais localement construit en fonction des caractéristiques de l'action et des conceptions des partenaires.

La question mérite cependant une attention particulière : dans la mesure où chaque évaluation est contextualisée, où elle concerne des actions et des acteurs situés, et où ses résultats doivent prendre sens pour ceux à qui ils sont communiqués, on peut s'interroger sur la capacité de généralisation induite par des outils trop particuliers (Dauvisis, 1995) et sur la difficulté de comparaison avec d'autres situations. Il ne faut cependant pas en déduire qu'une uniformité d'outils permettrait des conclusions plus assurées, mais rechercher à fonder les bases de comparaisons un tant soit peu valides. Le problème sous-jacent est celui du transfert implicite en sciences humaines du modèle de la mesure des sciences physiques et de sa validité pour les situations concernées : contrairement à ce que pourraient laisser croire la psychométrie et les modèles de la « vraie note » proposée par Piéron (1963) ou de courbe gaussienne si usuelle en notation scolaire, l'évaluation n'est pas une mesure (Noizet & Caverni, 1979), mais une pratique sociale ; aussi, les seuls modèles de la mesure les plus adaptés seraient ceux de l'économie et non ceux de la physique³.

Les outils d'évaluation eux-mêmes peuvent être objet d'évaluation, et leur choix répond à un certain nombre de critères déterminés par la finalisation de l'évaluation, la nature des données et les décisions à prendre : face à la diversité des pratiques d'ingénierie en évaluation, il y a là comme une méta-évaluation, véritable question pour la recherche qui, sans proposer de solution unique, pourrait au moins apporter l'éclairage d'une certaine rigueur scientifique avec des questionnement fertiles.

***Selon les qualités des outils,
quelle confiance méritent les résultats ?***

En logique, on sait que des conclusions élaborées à partir de prémisses fausses ne peuvent prétendre à aucune valeur de vérité ; par ailleurs, la solidité d'une chaîne correspond à celle de son maillon le plus faible : de même, la rigueur d'une démarche dépend de celle de chacune de ses étapes. Un outil ne disposant pas de qualités intrinsèques suffisantes met en péril la collecte de données pertinentes, valides et fiables (De Ketele & Roegiers, 1993) sur lesquelles doivent se fonder les résultats d'une évaluation.

Certaines évaluations institutionnelles n'échappent pas à de telles critiques, même lorsque leurs résultats bénéficient d'une large diffusion dans le public, leurs conclusions étant de fait souvent établies dans des conditions et sur des bases plus que critiquables (Vrignaud et coll., 2001). L'évaluation serait-elle une pratique si naturelle (Chardenet, 1999) qu'il suffirait d'être doué de raison pour « naturellement » produire des évaluations valides et dont les résultats soient fiables et généralisables ?

Quel que soit l'objet évalué, construire un outil d'évaluation suppose donc d'en connaître et d'en vérifier à la fois les qualités intrinsèques (Morissette, 1993), et l'adaptation aux finalités de l'évaluation (Cardinet, 1972; 1986a; 1986b): instrumenter l'évaluation l'exige – même en évaluation formative des apprentissages (Scallon, 2000) –, si la rigueur est effectivement recherchée pour des résultats ayant quelque fiabilité et pour des conclusions un minimum avérées. C'est pourquoi l'absence fréquente de ce préalable manifeste à l'évidence que, dépassant les exigences méthodologiques et pédagogiques, l'évaluation s'inscrit de fait dans des perspectives plus larges qui relèvent d'approches politiques et culturelles (Dauvisis, 1999 ; Dauvisis et al., 2003).

Rappeler les qualités attendues d'un outil d'évaluation et soumettre les pratiques d'évaluation et les conclusions à un tel examen devient une exigence déontologique pour les chercheurs en évaluation. Au Canada, plusieurs exemples récents témoignent d'une telle prise de responsabilité parmi les chercheurs (Joint Comitee, 1993 ; Séguin, Auger et al., 2001): il est vrai que leurs avis et leurs travaux bénéficient d'une meilleure reconnaissance sociale que ceux des chercheurs en éducation dans les milieux éducatifs français et que, par ailleurs, les partenaires du système éducatif n'hésitent pas à porter les problèmes d'injustice en matière d'évaluation scolaire devant les juridictions civiles (Lemay, 2000)!

Et si le roi était nu !

Évaluation : légitimité sociale ou rationalité ?

Démarche de preuve – ou au moins souvent présentée comme telle – l'évaluation est en réalité une pratique sociale contextualisée, régie par des normes, des coutumes et des valeurs qui ne s'inscrivent pas toutes dans la logique d'une totale rationalité. Comme toute pratique sociale, elle n'échappe pas à des jeux d'acteurs nombreux, et l'implicite y règne de manière évidente permettant aux différents acteurs de détenir une part de pouvoir à préserver jalousement: toute tentative de clarification est perçue avec un risque de mise en péril d'un équilibre social toujours fragile.

La recherche se trouve ainsi amenée à remettre en question la réalité associée à ce type d'activité.

- Des résultats établis en docimologie montrent depuis plus de huit décennies le peu de fidélité des questions ouvertes de rédaction: pourquoi celles-ci demeurent-elles les modalités privilégiées pour l'évaluation des connaissances dans notre système français? pourquoi les mesures de

modération *a priori* sont-elles si peu utilisées ? pourquoi les examinateurs, membres de jurys souverains, ne sont-ils souvent pas même informés de ces résultats ? pourquoi ces acquis ne font pas obligatoirement partie de la formation professionnelle des enseignants ? (Jorro, 2000.)

- Pourquoi les outils utilisés par les formateurs ou les évaluateurs ne sont pas systématiquement présentés avec les résultats obtenus ? Par exemple, dans un conseil de classe, chaque élève est distingué par une note par domaine d'enseignement, mais nul ne dit généralement comment cette note est obtenue : les types de situations d'évaluation, leur nombre et leur variété, les critères prioritaires, etc. Dans une évaluation institutionnelle, ce sont aussi les seuls résultats qui intéressent : par exemple, lorsque les médias se font l'écho d'un analphabétisme important à l'entrée en secondaire ou dans la population, rares sont les auditeurs qui cherchent à savoir comment et sur quelles compétences en lecture de telles conclusions sont avancées, ni même si la définition de l'analphabétisme, support de l'étude, est celle qu'ils interprètent.
- Certains prônent l'autoévaluation, mais que recèle de fait cette approche ? Le discours humaniste qui la soutient, comment s'inscrit-il dans un projet de sélection scolaire ou seulement de régulation sociale ? N'y a-t-il pas illusion lorsque l'autoévaluation se confronte à la certification ou prétend y contribuer ? (Depover & Noël, 1999, chap. 2.)
- Admettre un regard extérieur sur ses pratiques d'évaluation et exposer ses outils, n'est-ce pas, d'une certaine manière, s'exposer soi-même, au risque d'être dépouillé d'une carapace protectrice, d'être mis en cause dans ses valeurs personnelles de références et dans l'intime de ses convictions, et de perdre une part de pouvoir discret ou discrétionnaire ? De même que tous les élèves ne sont pas toujours prêts à exhiber leurs erreurs, même si l'enseignant tente de les convaincre de l'importance de l'approche formatrice de l'évaluation pour leur réussite, les enseignants craignent de se mettre à nu au travers de leurs pratiques d'évaluation : lorsqu'un regard externe vient institutionnellement observer ces pratiques, leur identité professionnelle elle-même peut en être transformée (Dauvisis, Droyer & Frossard, 1996).
- Enfin, en observant les pratiques mises en œuvre dans divers pays, on en vient aussi à se demander pourquoi les modalités d'évaluation et l'exigence de rigueur dans la construction des outils varient-elles si fortement selon les pays.

Au travers de ces quelques questions transparaissent fortement les aspects sociologiques et culturels de l'évaluation, qui se révèle comme une pratique sociale et qui, notamment dans l'usage et l'élaboration de ses outils, obéit à des logiques imposées par le contexte plus qu'à une rationalité scientifique délibérée et manifeste, comme l'indiquait le titre de l'ouvrage de Latour en 1992, «les réseaux que la raison ignore». La recherche en évaluation en prend désormais acte et, sans mettre de côté les préoccupations méthodologiques, focalise des travaux sur la précision de ces enjeux latents et des systèmes de valeurs qui régissent plus ou moins ouvertement l'activité évaluative.

***Mesure ou méthodologie :
quel modèle de vérité pour l'évaluation ?***

Parmi les débats culturels qui traversent l'évaluation, il en est un dont témoigne la signification même du sigle ADMEE des deux côtés de l'Atlantique : le «M» signifie ici «méthodologie», là-bas «mesure». Aucun des chercheurs de cette association ne limite l'évaluation à une mesure, au sens physique du terme, et la part de jugement dans cette activité humaine n'est niée par personne ; cependant le travail sur les qualités métrologiques des instruments, souvent ignoré de ce côté-ci, fait partie des préoccupations majeures de nos collègues d'outre-Atlantique. Il n'existe, notamment en France, aucun département d'évaluation au sein des universités ; au Canada, chacune des principales facultés d'éducation comporte un département «Mesure et évaluation» et dans toutes les recherches, la validation des instruments constitue une étape préliminaire fondamentale de la «cueillette» des données. On peut être surpris de ces différences ; on serait, en effet, enclin à penser que la rigueur scientifique cartésienne prédomine ici sur le pragmatisme ambiant qui tend à imprégner la culture anglophone et nord-américaine : force est de constater que c'est l'inverse et que les recherches et les dispositifs qui visent à «assurer la mesure» en évaluation (Cardinet & Tourneur, 1985 ; Cardinet, 1986c) ont encore un bien faible écho dans les pratiques européennes, voire sont totalement ignorés par la plupart des évaluateurs. Insister sur la méthodologie plutôt que sur la mesure relève sans doute d'un débat culturel, mais risque aussi d'occulter la méthodologie au profit de la seule production de résultats, parfois peu crédibles : rares étaient les chercheurs de l'ADMEE-Europe qui y portaient attention⁴ ; quelques autres commencent désormais à y être sensibles. Les enquêtes internationales et leurs résultats de même que les nouveaux supports d'évaluation créent désormais aussi en Europe de nouveaux questionnements méthodologiques.

Nouvelles questions : nouveaux outils et nouveaux supports

Les évolutions technologiques récentes dont l'usage se développe rapidement dans l'univers de l'éducation, peuvent, pour l'évaluation des apprentissages, être considérées sous deux aspects particuliers qui, respectivement, amènent à repenser les situations d'évaluation et leurs objets et permettent de développer une instrumentation sur de nouveaux supports, rapidement disponibles pour tous.

- Le développement d'Internet modifie les conditions d'accès des élèves à l'information et diversifie les sources de savoirs; par là-même s'introduisent des suspicions quant à l'origine personnelle des productions présentées et au degré d'appropriation des apprentissages correspondants, au point que les médias nationaux se font désormais l'écho, dans le grand public, du désarroi des enseignants quant à la fiabilité des réponses obtenues. L'urgence se fait sentir de repenser la nature des situations traditionnelles d'évaluation scolaire, dans ce nouveau contexte.
- Les développements technologiques au service de l'éducation enrichissent les possibilités pour l'instrumentalisation de l'évaluation, en créant des banques d'items et des procédures automatisées plus ou moins sophistiquées – ingénierie, rapidement objet d'un commerce rentable –, utilisées pour l'autoévaluation en autoformation et en enseignement individualisé, le suivi individuel ou l'usage collectif en salle de classe ou en enseignement à distance, et l'exploitation de tests standardisés pour des évaluations conçues sur le plan national ou international. Là encore l'ingénierie met en question la recherche, et de manière très urgente compte tenu de la rapidité des développements et de l'implantation des NTICE; jusqu'alors, en Europe francophone, rares étaient les équipes universitaires travaillant dans ces domaines et, il faut le reconnaître, l'intérêt pour leurs publications est souvent assez limité (Depover & Srebelle, 1996; Leclercq et al., 1998); c'est surtout dans les départements Mesure et évaluation du Canada que des travaux en ces domaines se sont le plus développés. Là encore, le contexte culturel imprime ses exigences à l'évaluation et à son instrumentation: cependant, une nouvelle ère semble se dessiner qui pourrait se jouer des frontières!

L'évaluateur : un ingénieur, un expert ou un bricoleur ?

C'est l'image professionnelle de l'évaluateur qui transparait au travers de l'instrumentation et de l'usage des outils. Un ingénieur est un concepteur qui propose un projet intégrant des données scientifiques et des savoirs stabilisés pour prévoir et organiser une réponse technique à un problème donné, en prenant en compte les différentes variables du contexte. Un expert est celui qui, en situation, pose un diagnostic et trouve une réponse pertinente ; pour cela, il dispose d'une instrumentation riche et parfaitement adaptée dont il se sait se servir à bon escient : chaque outil est totalement déterminé par sa fonctionnalité et aucune substitution n'est possible. Le bricoleur, quant à lui, fait avec ce qu'il a ou ce qu'il trouve et ne connaît pas toujours l'usage précis de l'outil qu'il utilise : « il fait avec » ou s'arrange, au risque de casser l'outil ou de transformer son projet ; certes, il y a des bricoleurs de génie, reconnus pour leur ingéniosité, mais, en général, évoquer l'image du bricoleur correspond plutôt à décrire un non-professionnel.

Tandis que la recherche en éducation et les institutions de formation tendent à présenter l'enseignant comme professionnel (Paquay et al., 1996), c'est l'image du bricoleur qui est communément répandue parmi les praticiens, ne nécessitant pas de formation particulière puisqu'il faut « se débrouiller » en situation. Ainsi, dans leur fonction d'évaluateur, les enseignants n'éprouvent guère de besoin de formation, ni ne s'inquiètent des acquis de la recherche : l'habitude et les manuels scolaires leur fournissent des outils d'évaluation qu'ils bricolent, sans souci de leur validité et de leurs effets sur les réponses. L'habitude crée des normes que nul ne songe à remettre en cause ou même à remettre en question.

Pour ce qui est des évaluations de projets ou de programme, la commande consiste à fournir des résultats, des évaluateurs sont sollicités ; dans la mesure où l'activité évaluative se présente comme une fonction quasi naturelle de tout être humain (Chardenet, 1999), chacun peut se déclarer compétent, notamment quand la culture d'évaluation portée par les pratiques ne se fonde, de fait, que sur les enjeux sociaux vécus par les divers partenaires. Tout questionnement porté par la recherche devient quasi inconvenant dans la mesure où il est susceptible de remettre en cause la nature et la valeur des résultats produits, et par là-même, la compétence et l'image sociale de ceux qui les produisent. L'engagement personnel des évaluateurs et leur bonne foi ne fait souvent pas de doute, mais des questions essentielles demeurent : les systèmes de valeurs et les références sur lesquels ils s'appuient dans leur activité

d'évaluation sont-ils vraiment aussi universellement partagés qu'ils peuvent le penser? où et comment valider ce postulat d'éventuelle universalité des valeurs? est-ce tant soit peu envisageable d'exiger une explicitation des modèles sous-jacents?

On le voit, l'ingénierie développe une instrumentation multiforme et souvent totalement empirique et non contrôlée; l'urgence devient manifeste d'un développement particulier des recherches en évaluation au sein de la recherche en éducation, pour espérer sortir de cet actuel brouillard ambiant, si tant est que le désir de rationalité soit effectivement une des valeurs acceptée pour l'évaluation: c'est aujourd'hui une des questions vives de la recherche (Dauvisis, 2000), qui demeure une véritable question ouverte tant les enjeux dépassent le cadre strict de la logique de la production de connaissances, pour s'inscrire dans une logique du politique et du social.

NOTES

1. Ce travail est en cours et une rubrique «terminologie» est désormais systématiquement présente dans chacun des bulletins de l'ADMEE-Europe.
2. Titre du Colloque de l'ADMEE-Canada en 1997, à Hull, et de la revue *Mesure et évaluation en éducation*, 19(2) 1996.
3. Cf. La rubrique «Mesure» de l'*Encyclopaedia Universalis*, qui présente la mesure dans trois domaines: la physique, la musique et l'économie.
4. Ainsi parmi les 40 articles publiés depuis 20 ans dans *Mesure et évaluation en éducation* (ADMEE, 2000), concernant des «instruments d'évaluation», seuls cinq proviennent d'auteurs européens: trois d'auteurs suisses (dont deux de Cardinet et un de Tschoumy, IRDP) et deux d'auteurs belges (Tourneur, Université de Mons, et Huynen, UCL).

RÉFÉRENCES

- Abernot, Y. (1988). *Les méthodes d'évaluation scolaire*. Paris: Bordas.
- ADMEE (2000). *Mesure et évaluation en éducation: 1978-1998*. (CD-ROM diffusé par l'ADMEE-Europe).
- Altet, M. (1994). *La formation professionnelle des enseignants*. Paris: PUF.
- Ardoino, J., & Berger, G. (1986). L'évaluation comme interprétation. *Pour*, 107, 120-127.
- Bonniol, J.J., & Vial, M. (1997). *Les modèles de l'évaluation, textes fondateurs avec commentaires*. Bruxelles: De Boeck Université.
- Bressoux, P., Leroy-Audouin, C., & Coustère, P. (1998). Les extensions des modèles multiniveaux et leur application pour l'évaluation en éducation. *Mesure et évaluation en éducation*, 21(1), 39-61.
- Bloom, B.S., et al. (1969). *Taxonomie des objectifs pédagogiques: I. Domaine cognitif* (trad. Lavallée). Montréal: Éditions nouvelle.

- Cardinet, J. (1972). *L'adaptation des tests aux finalités de l'évaluation*. Neuchâtel : IRDP.
- Cardinet, J. (1986a). *Pour apprécier le travail des élèves*. Bruxelles : De Boeck Université.
- Cardinet, J. (1986b). *Évaluation scolaire et pratique*. Bruxelles : De Boeck Université.
- Cardinet, J. (1986c). *Évaluation scolaire et mesure*. Bruxelles : De Boeck Université.
- Cardinet, J., & Tourneur, Y. (1985). *Assurer la mesure. Guide pour les études de généralisabilité*. Bern : Peter Lang.
- Casalfiore, S., & Paquay, L. (1998). *Le paradigme de l'action située: regard sur l'activité enseignante et implications pour l'évaluation de la qualification des enseignants*. Louvain-la-Neuve : UCL, Département des sciences de l'éducation.
- Cauterman, M.M., Demailly, L., et al. (1999). *La formation continue des enseignants est-elle utile ?* Paris : PUF.
- Chardenet, P. (1999). *De l'activité évaluative à l'acte d'évaluation*. Paris : L'Harmattan.
- Charlier, B. (1998). *Apprendre et changer sa pratique d'enseignement*. Bruxelles : De Boeck Université.
- Crahay, M. (éd.) (1994). *Évaluation et analyse des établissements de formation*. Bruxelles : De Boeck Université.
- CRESAS (1986). *Les élèves face à l'évaluation: enquête auprès de 15 établissements de l'académie de Nantes*. Paris : INRP.
- Cros, F. (dir.) (2000). *Le transfert des innovations scolaires. Une question de traduction*. Paris : INRP, Mission « Innovation et Recherche » (ronéo).
- Dauvisis, M.C. (1995). Analyse critique du livre de G. Figari «Évaluer: quel référentiel?». *Revue française de pédagogie*, 113, 144-146.
- Dauvisis, M.C. (1999). Valeurs latentes des pratiques de certification. *Spirale*, 23, 231-241.
- Dauvisis, M.C. (2000). *Introduction de l'atelier: « Rationalisation de l'action et évaluation: un problème culturel ? »* Cinquième Biennale de l'éducation et de la formation, Paris, Sorbonne. Actes de la cinquième Biennale (en ligne sur le site: [www.INRP.fr]).
- Dauvisis, M.C. (2001). *Outils et méthodes d'évaluation: quand l'ingénierie sollicite ou questionne la recherche*. Communication au Congrès d'actualité de la recherche en éducation et formation, AECSE, Lille (CD-ROM des Actes publié en 2003).
- Dauvisis, M.C. (2003). Les enseignants peuvent-ils croire à la formation? In G. Boutin (dir.), *La formation des enseignants en question. Modalités, entrée dans le métier et dimensions critiques* (pp. 159-190). Montréal : Éditions nouvelles.
- Dauvisis, M.C., Droyer, N., & Frossard, G. (1996). *Pratiques de CCF en BEPA*. Dijon : Educagri.
- Dauvisis, M.C., Millot, J., Frossard, G., & Droyer, N. (2003). *L'équité, un idéal: mais quelle réalité dans les pratiques d'évaluation dans les Bac?* Communication au Colloque de l'ADMEE-Europe: «L'évaluation entre équité et efficacité», Liège (à paraître au CD-ROM des Actes).
- De Ketele, J.M. (1993). L'évaluation conjugquée en paradigmes. *Revue française de pédagogie*, 103, 59-80.
- De Ketele, J.M., & Roegiers, X. (1993). *Méthodologie du recueil d'information, fondements des méthodes d'observation, les questionnaires d'interviews et d'études de documents*. Bruxelles : De Boeck Université.
- De Landsheere, G. (1974). *Évaluation continue et examens*. Précis de docimologie. Bruxelles-Paris : Labor-Nathan.

- De Landsheere, G. (1979). *Dictionnaire de l'évaluation et de la recherche en éducation*. Paris : PUF.
- Delory, C. (1996). Une épreuve de fin d'études primaires : levier pédagogique ou frein d'innovation. *Mesure et évaluation en éducation*, 18(3), 43-58.
- Depover, C., & Noël, B. (éd.) (1999). *L'évaluation des compétences et des processus cognitifs. Modèles, pratiques et contextes*. Bruxelles : De Boeck Université.
- Depover, C., & Srebelle, A. (1996). Fondements d'un modèle d'intégration des activités liées au NTIC dans les pratiques éducatives. In L.C. Pochon & A. Blanchet (éd.), *Information et éducation : regards cognitifs, pédagogiques et sociaux*. Paris : INRP, 9-20.
- Dominice, P. (1979). *La formation, enjeu de l'évaluation*. Berne : Peter Lang.
- Figari, G. (1994). *Évaluer : quel référentiel?* Bruxelles : De Boeck Université.
- Figari, G., & Achouche, M. (2001). *L'activité évaluative réinterrogée. Regards scolaires et socioprofessionnels*. Bruxelles : De Boeck Université.
- Hadji, C. (1997). *L'évaluation démystifiée*. Paris : ESF.
- Jarousse, J.P., & Piquée, C. (2001). *Évaluer l'accompagnement scolaire : comment la pratique de formation agit sur la pratique d'évaluation*. Communication au Colloque international de l'ADMEE-Europe, Aix-en-Provence. CD-ROM des Actes (atelier «Méthodes et outils»).
- Joint Comitee (1993). *Principes d'équité relatifs aux pratiques d'évaluation des apprentissages scolaires au Canada*. University of Alberta: Center for research in applied measurement and evaluation.
- Jorro, A. (2000). *L'enseignant et l'évaluation*. Bruxelles : De Boeck Université.
- Latour, B. (1992). *Les réseaux que la raison ignore*. Paris : L'Harmattan.
- Latour, B. (1995). *La science en action*. Paris : Gallimard.
- Leclercq, D., et al. (1998). L'amphithéâtre électronique. Une application de LORT-SAFE. In D. Leclercq (éd.), *Pour une pédagogie universitaire de qualité* (pp. 161-186). Spirmont : Mardaga.
- Lecoïnte, M. (1997). *Les enjeux de l'évaluation*. Paris : L'Harmattan.
- Legendre, R. (1993). *Dictionnaire actuel de l'éducation* (2^e éd.). Montréal : Guérin.
- Lemay, V. (2000). *Évaluation scolaire et justice sociale*. Montréal : ERPI.
- Monteil, J.M. (1989). *Éduquer et former. Perspectives psycho-sociales*. Grenoble : PUG.
- Morissette, D. (1993). *Les examens de rendement scolaire*. Québec : Presses de l'Université Laval.
- Nadeau, M.A. (1984). *L'évaluation de programme. Théorie et pratique* (2^e éd., 1988). Québec : PUL.
- Noizet, G., & Caverni, J.P. (1979). *Psychologie de l'évaluation scolaire*. Paris : PUF.
- Paquay, L. (1994). Vers un référentiel des compétences professionnelles de l'enseignant. *Recherche et formation*, 16, 7-33.
- Paquay, L., et al. (éds) (1996). *Former des enseignants professionnels : quelles stratégies ? quelles compétences ?* Bruxelles : De Boeck Université.
- Peretti, A. de (1998). *Encyclopédie de l'évaluation et de la formation en éducation. Guide pratique*. Paris : ESF.
- Piéron, H. (1963). *Examens et docimologie*. Paris : PUF.

- Rayou, P. (2001). *Les lycéens dissertateurs*. Communication au Congrès de l'AECSE, Lille (publiée au CD-ROM des Actes en 2003).
- Scallon, G. (2000). *L'évaluation formative*. Bruxelles: De Boeck Université.
- Schön, D. A. (1983) *The reflective practitioner*. New York: Basic Books. (Trad. française: *Le praticien réflexif. À la recherche du savoir caché dans l'agir professionnel*. Montréal: Éditions Logiques.)
- Séguin, S.P., Auger, R., et al. (2001). *Pour une politique des apprentissages scientifiquement fondée. Avis d'experts en mesure et évaluation des apprentissages sur le projet de Politique d'évaluation des apprentissages*. Montréal: LABFORM, Faculté d'éducation, UQAM.
- Solaux, G. (éd.) (2000). *L'évaluation des politiques d'éducation*. Dijon: CRDP. (Coll. Documents et rapports pour l'éducation)
- Stufflebeam, D., et al. (1980). *L'évaluation en éducation et la prise de décision*. Ottawa: NHP.
- Vial, M. (2001). *Se former pour évaluer. Se donner une problématique et élaborer des concepts*. Bruxelles: De Boeck Université.
- Vrignaud et coll. (2001). *Rapport sur les analyses et études conduites par l'INETOP/CNAM, la demande du HCEE sur « Que savons-nous des connaissances des élèves de fin de collèges », les évaluations de la DEP/DPD*. Paris: INETOP, ministère de l'Éducation nationale.
- Weiss, J. (1994). Évaluer autrement. *Mesure et évaluation en éducation*, 17(1), 63-73.
- Weiss, J. (éd.) (1991). *L'évaluation: problème de communication*. Fribourg: DelVal.