

Dix années de travaux européens et nord-américains sur l'évaluation : quelles lignes de force ?

Jean Cardinet and Dany Laveault

Volume 18, Number 3, 1996

URI: <https://id.erudit.org/iderudit/1092252ar>

DOI: <https://doi.org/10.7202/1092252ar>

[See table of contents](#)

Publisher(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (print)

2368-2000 (digital)

[Explore this journal](#)

Cite this article

Cardinet, J. & Laveault, D. (1996). Dix années de travaux européens et nord-américains sur l'évaluation : quelles lignes de force ? *Mesure et évaluation en éducation*, 18(3), 1-25. <https://doi.org/10.7202/1092252ar>

Article abstract

A review of literature covering the last 10 years of European and North American research efforts reveals the numerous progress that have been accomplished in educational evaluation. The authors point to the main development which occurred in relation to the three poles of the instructional triangle (learner, teacher, subject matter) as well as to a fourth one, society, which they added to the previous three. They acknowledge the increasing role played by cognitive sciences in educational evaluation and the greater complexity of the evaluation task. As a conclusion, they isolate from their own observations a series of focal directions for the near future.

Dix années de travaux européens et nord-américains sur l'évaluation: quelles lignes de force?

Jean Cardinet
Neufchâtel

Dany Laveault
Université d'Ottawa

Un état de la question des dix dernières années de travaux européens et nord-américains révèle que de nombreux progrès ont été réalisés en matière d'évaluation en éducation. Les auteurs font ressortir les principaux développements qui se sont produits en rapport avec les trois pôles du triangle didactique (apprenant, enseignant, savoir), auxquels ils ont ajouté un quatrième pôle, celui de la société. Ils constatent l'importance accrue des sciences cognitives dans l'évaluation en éducation et la complexité grandissante de la tâche d'évaluer. En conclusion, ils dégagent de leurs observations quelques lignes de force pour l'avenir immédiat.

(évaluation pédagogique, processus cognitifs, didactique, méthodes d'évaluation, histoire de l'évaluation pédagogique, système d'évaluation, pratique évaluative)

A review of literature covering the last 10 years of European and North American research efforts reveals the numerous progress that have been accomplished in educational evaluation. The authors point to the main development which occurred in relation to the three poles of the instructional triangle (learner, teacher, subject matter) as well as to a fourth one, society, which they added to the previous three. They acknowledge the increasing role played by cognitive sciences in educational evaluation and the greater complexity of the evaluation task. As a conclusion, they isolate from their own observations a series of focal directions for the near future.

(educational evaluation, cognitive processes, didactics, evaluation methods, history of educational evaluation, evaluation system, evaluative practice)

Texte de la conférence d'ouverture présentée par les deux auteurs au 10^e colloque annuel de l'Association pour le développement de la mesure et de l'évaluation (ADMÉE-Europe), tenu à Grenoble le 18 septembre 1996. Ce texte est publié sur invitation aux auteurs par la Revue *Mesure et évaluation en éducation* (volume 18, n° 3) et reproduit dans Marsyas avec son autorisation. Tous droits réservés: Revue *Mesure et évaluation en éducation*.

L'évaluation, étude de modèles

L'évaluation est un domaine de recherche, qui a besoin d'expérience et de théorie, mais ce n'est pas une science qui explore un réel extérieur à nous, comme la physique ou la sociologie. Elle doit plutôt créer des modèles et essayer de les appliquer, comme le font la statistique ou la recherche opérationnelle. Son premier cadre de référence est la théorie de la décision. Voilà pourquoi la méthodologie de l'évaluation peut progresser aussi bien à partir des demandes de la pratique (parce qu'un besoin existe auquel il faut que des spécialistes répondent), qu'à partir de l'offre de nouveaux modèles (parce que des disciplines proprement scientifiques ont découvert des relations qui peuvent éclairer cette méthodologie).

Du point de vue de la pratique, on peut dire que ces dernières années ont vu le triomphe de la culture de l'évaluation. Dans tous les domaines, on s'est efforcé de contrôler l'effet des décisions prises pour diriger rationnellement les projets et les institutions. Cette nouvelle philosophie politique a été immédiatement appliquée à la gestion des innovations (par exemple à l'introduction de nouveaux programmes d'études, ou bien à la fixation d'exigences sur le plan de la redevabilité).

L'évaluation des élèves, souci presque exclusif à l'origine, n'est plus ainsi qu'un problème d'évaluation parmi d'autres, même si, en raison de la multiplicité des travaux conduits sur ce thème, la majeure partie de cet exposé devra y être consacrée. Finalement, on accepte aujourd'hui qu'au lieu d'être le couperet qui tranche définitivement le sort des élèves, l'évaluation devienne l'instrument même de leur progrès.

Du point de vue de la recherche fondamentale, c'est surtout le triomphe de l'approche cognitive qui a marqué la décennie, et en particulier le développement des conceptions néo-constructivistes, qui présentent l'apprentissage comme le produit d'un processus de construction sociale. Ces idées ont influencé les méthodes de recherche de toutes les sciences sociales. Dans tous les domaines, l'étude des stratégies des acteurs (et de leurs négociations sociales) s'est révélée primordiale, entraînant l'adoption par les chercheurs d'une nouvelle épistémologie, interactionniste. On verra que cette approche renouvelle fondamentalement la problématique de l'évaluation, en mettant en cause l'idée même d'un résultat objectif en matière d'apprentissage.

Le sujet est si vaste et l'espace imparti dans une revue si court que nous ne pourrions qu'évoquer brièvement chaque domaine. Nous avons accepté ce risque de superficialité pour mettre en évidence l'étendue surprenante du

champ que l'évaluation a investi durant la dernière décennie. Elle touche tous les domaines des sciences sociales et est elle-même concernée par tous leurs développements.

Pour classer la multitude d'études qu'il nous fallait aborder, nous avons choisi le fameux triangle didactique, reliant l'apprenant, l'enseignant et le savoir. À ces trois pôles nous en avons ajouté un quatrième, celui de la société, pour intégrer les recherches des sociologues de l'éducation, qui jettent un jour nouveau sur tous les problèmes pédagogiques. Le schéma organisateur qui a servi à structurer notre texte est illustré dans la figure 1.

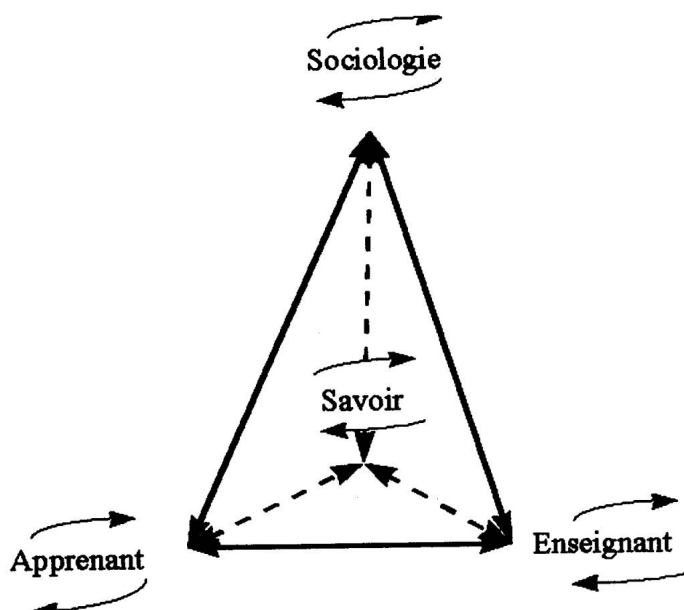


Figure 1- Les quatre pôles de la recherche en évaluation

Les progrès accomplis

Le pôle de l'Apprenant

La compréhension que l'élève a du savoir

Nouvelles pratiques dans la formulation des objectifs. - La dernière décennie a conduit à redéfinir le sens d'une pédagogie et d'une évaluation par objectifs.

La définition des objectifs, en effet, est de plus en plus remise en question dans sa formulation comportementale habituelle. L'élève et le maître s'y perdent lorsque les objectifs sont trop nombreux et le savoir ainsi morcelé.

Les nouveaux objectifs décrivent davantage des résultats globaux, des compétences appliquées à des projets intégrateurs et signifiants pour l'élève. L'objectif «jargon de spécialiste» (Conseil supérieur de l'éducation, 1992) que l'on a retrouvé dans certains bulletins critériés au Québec cède maintenant la place à un objectif décrit sous l'angle de progrès individuels précis, inséré dans un contexte psychosociologique réaliste et débouchant sur des développements généraux de la personne.

Les recherches de psychologie cognitive. - Une telle évolution résulte sans aucun doute des travaux de la psychologie cognitive. Ceux-ci ont bien montré qu'un élève pouvait répondre juste pour une raison erronée et que sa démarche de pensée était plus importante que son résultat. Le maître doit comprendre les réponses avant de les mesurer!

Toute une série de recherches, en particulier dans le domaine des connaissances scientifiques (par exemple Carraher, Carraher et Schliemann 1985; Viennot, 1979;) ont révélé l'existence de champs de représentations parallèles. Les apprentissages scolaires se juxtaposent aux savoirs intuitifs antérieurs, sans les remplacer, de sorte qu'une fois hors de l'école, les gens se comportent comme s'ils n'avaient rien appris, alors qu'ils savent répondre correctement aux examens.

Les implications pour une didactique et une évaluation individualisées sont évidentes, mais restent encore à développer pratiquement. Il faut poser aux apprenants des problèmes qui suscitent un conflit entre leurs représentations naïves et leurs savoirs élaborés, pour qu'ils soient amenés à restructurer l'ensemble de leurs connaissances.

La compréhension que l'élève a des demandes du maître

La communication des objectifs en pratique. - L'objectif, en tant qu'outil de communication entre le maître et l'élève, peut, tout comme l'enseignement, être partagé, négocié ou simplement transmis, selon l'approche pédagogique privilégiée par le professeur. Diverses tentatives innovatrices ont montré qu'une meilleure connaissance des objectifs par l'élève lui permettait de mieux planifier son travail d'études, en s'organisant en fonction d'attentes précises (Doyon, 1991).

Nunziati (1990), s'inspirant de Bonniol (1981), a fondé toute une didactique sur l'appropriation par l'apprenant des critères d'évaluation, à partir de

l'analyse de travaux d'élèves déjà corrigés par le professeur, puis de travaux de camarades discutés en groupes. L'étudiant devient ainsi capable de se situer lui-même par rapport aux objectifs pédagogiques et de gérer son propre apprentissage.

Recherches sur la construction de l'intersubjectivité.- Les recherches de psychologie sociale cognitive qui ont analysé de près le dialogue maître-élève ont montré que la compréhension réciproque se développait comme une construction conjointe (Grossen, 1988, Perret-Clermont et Nicolet, 1988), où chacun essayait de lire la pensée de l'autre, avec tous les processus de formulation d'hypothèses et de vérification que comporte la lecture. Il en résulte que l'apprentissage est un processus dynamique, et qu'on le fausse lorsqu'on veut en faire un bilan soi-disant «objectif».

Le résultat de l'élève dépend notamment du contexte social dans lequel le problème est posé, et en fonction duquel l'élève interprète la question (Schubauer-Leoni, Bell, Grossen et Perret-Clermont, 1989). On ne peut donc pas mesurer son niveau de connaissances de façon «absolue» (Cardinet, 1989). Sa réussite est une fonction à deux variables, où l'examineur est aussi déterminant que l'examiné. Ceci remet en cause la signification véritable des examens, en guidant les recherches vers des formes de certification plus proches de la vie réelle.

La compréhension que l'élève a de lui-même, apprenant

Les pratiques d'auto-évaluation.- Comme il est difficile de fournir un *feedback* individuel à chaque élève dans une situation d'enseignement collectif, il a fallu développer des pratiques d'auto-évaluation et de coévaluation octroyant à l'élève une plus grande part de contrôle sur la régulation de ses apprentissages et de ceux de ses pairs (Darmuzey, 1988; Scallon, 1988; Sevikian, 1986).

Pour y parvenir, il a fallu revoir tout le processus de transmission des notes du maître aux élèves et, à travers eux, aux parents. Ceci s'est traduit par de nombreux efforts, notamment au Québec, pour que le bulletin critérié soit le reflet des pratiques d'enseignement et un outil de transmission des résultats compréhensible par les destinataires (Conseil supérieur de l'éducation, 1992).

Au primaire, la présentation du bulletin aux parents est devenue l'occasion pour l'élève de faire un retour sur le travail accompli, d'explicitier ses acquisitions et de se fixer de nouveaux objectifs. C'est aussi l'occasion d'un dialogue entre l'élève, les parents et l'éducateur sur les moyens d'atteindre ces objectifs (Doyon, 1991).

Les recherches sur la métacognition et la motivation scolaire. - Sur le plan métacognitif, l'auto-évaluation encourage la pensée critique, les apprentissages de type réflexif et la prise en charge par l'élève de ses processus cognitifs (Noël, 1991), notamment de tout ce qui concerne ses démarches de mémorisation ou de correction d'erreurs (Romainville, 1993; Vermersch, 1994).

La psychologie dynamique du développement met aussi en évidence que l'élève, mieux informé des objectifs, des attentes et des exigences à son égard, anticipe mieux ce que l'on attend de lui (Fournier et Laveault, 1994; Laveault et Fournier, 1990; Wolfs, 1993). Des objectifs rapprochés et la perception du but à atteindre développent un sentiment d'efficacité personnelle, qui contribue à l'apparition de la motivation intrinsèque (Bandura, 1993; Scallon, 1992).

Le pôle de l'Enseignant

La compréhension que le maître a du savoir

La demande pratique de critères de réussite. - Les nouveaux programmes d'études présentent de nouveaux défis aux enseignants en matière de formulation d'exigences. Les demandes accrues en matière de redevabilité obligent les enseignants à se concerter afin de déterminer les standards minimaux de réussite. C'est le cas de plusieurs examens de certification professionnelle aux États-Unis et de l'épreuve synthèse de programme dans les Collèges du Québec (Ministère de l'Éducation du Québec, 1993).

La description d'objectifs d'intégration dans les nouveaux programmes d'études soulève aussi plusieurs difficultés. Ces objets d'évaluation ne peuvent être évalués qu'en situation de «mesure authentique». Les enseignants doivent donc concevoir des situations qui placent l'évalué en situation réelle ou proche du réel. Les performances complexes réalisées dans de telles conditions doivent être analysées au moyen de critères observables, valides et fidèles (Bertrand, 1996).

Des exemples sont souvent plus parlants que des définitions. En France, en début de CE2, de 6^e et de Seconde, des évaluations des compétences et des connaissances des élèves sont effectuées chaque année sur un plan national. Elles visent entre autres choses à mettre à disposition des enseignants des repères pour ajuster leur niveau d'exigence. Elles sont même complétées depuis 1992 par d'autres instruments d'évaluation (MEN-DEP, 1996).

L'étude de la transposition didactique. - C'est aux didacticiens français que revient le mérite d'avoir analysé les processus complexes de transformation

d'un corps de savoirs savants, comme la mathématique, en savoirs enseignables à un niveau donné de la scolarité (Chevallard, 1985). Les professeurs d'école et de collège constituent des maillons intermédiaires dans cette chaîne de retraductions, dont la finalité est bien souvent le contrôle et le classement des élèves. Ces retraductions sont sources d'erreur dans l'évaluation, dans la mesure où elles mettent les élèves en face d'exigences différentes d'une classe à l'autre.

La façon même de formuler les questions relatives à un contenu donné comporte aussi une large part d'arbitraire, car la bonne réponse doit être en partie cachée, pour qu'il reste un problème. Les élèves doivent donc décoder les intentions de leur professeur (Brousseau, 1986). L'inégalité devant l'évaluation résulte du degré variable de connivence avec l'enseignant, dont bénéficient les enfants, selon leur milieu social d'origine. Lorsque l'on autorise une telle connivence, l'on s'aperçoit que les élèves peuvent constituer d'excellents juges de la familiarité de la tâche à accomplir et assister l'enseignant dans l'«habillage» des situations didactiques (Jonnaert et Laveault, 1994).

La compréhension que le maître a de lui-même, évaluateur

La demande de formation à l'évaluation. - Dans tous les pays francophones, les sessions de formation à l'évaluation se sont multipliées ces dernières années pour permettre aux enseignants de repenser leurs pratiques. Il est indispensable, par exemple, que les évaluateurs s'entendent sur les critères, leur interprétation et leurs niveaux d'exigence (Eugène, Loncle Montméas, 1987; Lejong, Detheux de Bal, 1987).

Le travail de systématisation des démarches de l'évaluation, auquel les formateurs ont dû se livrer (par exemple ceux des MAFPEN, Missions Académiques à la Formation des Personnels de l'Éducation Nationale en France), a permis de clarifier la méthodologie de l'évaluation formative (Nunziati, 1989). On accepte désormais le fait qu'elle est subordonnée à la conception de la didactique choisie par l'enseignant (Bain et Schneuwly, 1993). En conséquence, l'idée d'instruments passe-partout, qui pourraient être utilisés sans que l'enseignant n'ait reçu une formation adéquate en matière d'évaluation, paraît désormais à abandonner.

Les recherches sur la démarche d'évaluation. - Passant à un niveau de généralité supérieur, les chercheurs en éducation, notamment les philosophes (Ardoino et Berger, 1989; Hadji, 1992), ont cherché à caractériser la démarche d'évaluation en elle-même. Ils ont montré qu'elle supposait certains choix épistémologiques préalables (Nadeau, 1988; De Ketele, 1993), qui expli-

quaient les désaccords possibles entre évaluateurs positivistes, naturalistes, ou interactionnistes (Cardinet, 1995).

Les méthodologues demandent désormais que toute la procédure soit négociée, depuis la mise au point de référentiels d'évaluation acceptés par les divers partenaires, jusqu'à la diffusion des résultats, qui doivent pouvoir susciter des boucles de remédiation (Black et Deakins, 1990).

La compréhension que le maître a des difficultés de l'élève

La demande de méthodes et d'instruments concrets. - Les enseignants, surtout du primaire et du secondaire, se sont appropriés graduellement les principes de base de l'évaluation formative. Cette assimilation, parfois déformante, s'est d'abord traduite par des modalités hybrides d'évaluation encore fortement teintées des pratiques antérieures d'évaluation normative: évaluation microsommative, sommative critériée, continue (au lieu de continue), «qui ne compte pas», etc. (Conseil supérieur de l'éducation, 1992).

Pourtant, la pertinence de l'évaluation formative est maintenant chose admise dans le corps enseignant. La difficulté est plutôt aujourd'hui de concrétiser cette intention généreuse. On peut relever avec plaisir la publication, ces dernières années, d'une série de livres écrits par des praticiens pour des praticiens (Groupe EVA, 1991; Guilbert, Hibou, Le Lièvre-Bourdin, Ouanas et Tavernier, 1995; Scallon, 1988; Veslin et Veslin, 1992). Leur but est bien de permettre aux professeurs de détecter les difficultés spécifiques de leurs élèves.

L'étape suivante serait maintenant de lier davantage évaluation formative et didactique de la branche. C'est là que les chercheurs en didactique vont avoir un rôle essentiel à jouer (Allal, 1991; Astolfi et Develay, 1989; Bain et Schneuwly, 1993; Giordan, 1989; Scallon, in Laveault (éd.), 1992; Wirthner, Martin, et Perrenoud (dir), 1991).

La recherche des obstacles épistémologiques. - Les travaux de recherche progressent en effet sur la genèse des notions mathématiques, scientifiques, etc. (Bednarz et Garnier, 1989; Giordan et De Vecchi, 1987; Audigier, 1987). Baliser un cheminement d'apprentissage précis pour l'élève apparaît bien illusoire, mais on peut repérer certaines difficultés fondamentales, qui font obstacle à l'intuition directe, et certaines hiérarchies dans la complexité des concepts, dont les didacticiens doivent nécessairement tenir compte (Astolfi, 1992).

Le cheminement de l'apprentissage de chaque élève dépend surtout de l'interaction entre ses conceptions et stratégies cognitives propres et les «perturbations» que l'enseignant peut introduire pour les faire évoluer (Giordan, 1995). L'autonomie de chaque élève, une fois reconnue, est renforcée par l'établissement de contrats d'apprentissage (Vial, 1987).

Le pôle du Savoir

La compréhension que le scientifique a du savoir

L'apparition de nouveaux savoirs pratiques. - Au cours des dernières années, les enseignants de plusieurs pays francophones ont vu apparaître de nouveaux objets d'évaluation dans les programmes d'étude officiels, par exemple dans les programmes intégrés en Belgique (CCEMPC, 1993). Sur le plan cognitif, il ne s'agit plus de notions isolées, ni d'objectifs comportementaux étroits, ni de capacités transversales abstraites, mais de plus en plus d'apprentissages intégrés, à la fois objectifs de transfert et de généralisation (Lussier-Charette, 1987). On ne veut plus tout évaluer, mais on veut évaluer «un tout», c'est-à-dire une compétence représentant une combinaison de capacités, applicable dans une situation concrète (Barth, 1993; Bronckart, 1985). Comme l'écrivent Eugène, Loncle et Montméas (1987, p. 17): «Ce ne sont pas les savoirs en eux-mêmes qui sont évalués, mais leur mobilisation pertinente en vue de la réalisation de la performance attendue.»

Sur ce plan pratique, les objectifs affectifs sont de plus en plus présents (Morissette, 1993; Morissette et Gingras, 1989). Au niveau primaire et au secondaire, ils jouent principalement une fonction instrumentale, asservie à l'atteinte des objectifs cognitifs: comment en effet enseigner les sciences sans favoriser des habitudes de travail rigoureuses, le goût de la découverte, etc.? Sur le plan de la formation professionnelle, les objectifs affectifs jouent un rôle plus délicat. Leur atteinte est parfois essentielle: par exemple, le respect du malade chez les infirmières. Mais comment contrôler s'ils ont été atteints? Comment les utiliser à des fins de sanction?

Les recherches sur l'expertise. - Ces changements quant aux objets d'évaluation ont été suscités par les résultats de recherches multiples, qui remettent en cause les façons antérieures de définir le savoir. Les études sur la mesure des acquis expérientiels (Barkatoolah, 1987; Meyer Berger, 1988), mais surtout les recherches opposant les processus mentaux des novices et des experts dans diverses professions (par exemple les enseignants: Tochon, 1993), ont montré que les savoirs et les savoir-faire isolés avaient un rôle relativement limité dans les compétences, que les stratégies d'attaque des problèmes étaient primordiales, mais en même temps qu'elles n'étaient pas formulables de façon explicite ni linéaire (Resnick et Ford, 1981).

On est donc loin de pouvoir cerner précisément la nature des savoirs à enseigner et les recherches devront se poursuivre dans ce domaine fondamental pour l'éducation.

La compréhension que le scientifique a de l'enseignant, évaluateur

La demande de certification. - Confrontés aux exigences sociales de certification, les enseignants s'efforcent de formuler des jugements, au mieux de leur conscience, mais toujours dans l'incertitude. C'est pourquoi de nouvelles procédures d'examen ont été mises au point ces dernières années, surtout dans le domaine professionnel (Montméas, Dauvisis, Chautard, Courtas et Frossard, 1989) pour contrôler des savoir-faire plus proches des exigences de la pratique, en observant les candidats dans des situations extra-scolaires. La publication préalable de référentiels contribue à fixer le niveau d'exigences (Eugène, Loncle et Montméas, 1987). Des systèmes de jurys et des contrôles statistiques ultérieurs permettent d'éviter certaines dérives possibles (Jaeger, 1994; Johnson, 1989).

Le remplacement des diplômes par des listes de compétences acquises est en discussion dans plusieurs systèmes de formation et est un fait accompli au Royaume-Uni. Liée d'abord à la certification des acquis expérimentiels, la technique des portfolios se répand maintenant en milieu scolaire (Paulson, 1996), pour remplacer les examens, dont le résultat est jugé trop peu valide (Paris et Ayres, 1994).

Les recherches sur l'évaluation. - De nombreux travaux ont pris comme objet d'étude la procédure d'évaluation elle-même. On contrôle le dispositif utilisé pour mesurer les performances (d'élèves, d'écoles, pour certains juges, etc.) lorsque cette performance est prise comme critère essentiel de réussite (Cronbach, Linn, Brennan et Haertel, 1995).

Le recours à l'évaluation fondée sur la performance peut révéler des effets d'ensemble importants (Grisay, 1988), mais soulève de nombreuses difficultés et devient très onéreux pour des décisions particulières, surtout lorsque les enjeux de l'évaluation sont élevés (Baker, O'Neil et Linn, 1993; Frechtling, 1991; Linn, 1994; Linn et Burton, 1994; Mehrens, 1992). C'est pourquoi, de plus en plus, on recherche de nouvelles formes d'évaluation en analysant les risques correspondants quant à la fiabilité des jugements d'experts (Reckase, 1995), à la validité des critères d'évaluation (Messick, 1989) et au coût financier que cela suppose.

Dans le domaine des programmes nationaux et internationaux d'évaluation, et en matière de *testing* adaptatif, le modèle de la Théorie de Réponse

aux Items (TRI) a permis de réaliser d'énormes progrès (Hambleton et Swaminathan, 1985; Hambleton, Swaminathan et Rogers, 1991). La mise à disposition de banques d'items précalibrés permet désormais de comparer les résultats d'échantillons différents d'items ou de sujets, par rapport au même trait latent (Bertrand et Jeanrie, 1994).

Plusieurs logiciels performants ont été mis au point, soit pour effectuer les études de généralisabilité: Etudgen (McNicoll, 1996) et GT (Ysewijn, 1996), soit pour calculer les paramètres d'items selon la TRI: BILOG2 (Mislevy et Bock, 1986).

La compréhension que le scientifique a de l'élève, apprenant

L'individualisation, dans la pratique du testing comme de l'enseignement. - Sur le plan de l'évaluation en salle de classe que des évaluations à grande échelle, on remet de plus en plus en question l'utilité de soumettre tous les évalués aux mêmes conditions d'évaluation. L'utilisation de la théorie de réponse aux items permet des «examens sur mesure» et simplifie la comparaison des individus à partir de leurs réponses à des questionnaires adaptatifs différents. Il existe d'amples démonstrations de la plus grande efficacité de telles approches pour la collecte de renseignements dans des situations de *testing* à grande échelle.

Le succès que rencontrent auprès des praticiens des ouvrages sur les styles cognitifs d'apprentissage des élèves révèle un autre besoin réel. Il ne suffit pas de pouvoir détecter les difficultés d'un étudiant: il faut savoir comment l'aider à les surmonter. C'est dans ce but qu'une meilleure connaissance des facteurs individuels affectant l'activité cognitive est souhaitable (La Garanderie, 1980; Bandler et Grinder, 1990).

Les recherches sur les particularités individuelles. - Ce sont les recherches en psychologie différentielle qui peuvent donner réponse à ces demandes d'une évaluation sur mesure, individualisée et adaptée aux besoins de l'élève.

La principale crainte que suscite chez les experts l'emploi du *testing* adaptatif est celle de voir les estimations de compétence faussées par des «biais sociaux» incontrôlés (Dechef et Laveault, 1993). Les recherches récentes sur le fonctionnement différentiel des items répondent à cette préoccupation d'équité envers les évalués. Il s'agit d'éviter que leur origine culturelle, ethnique, socio-économique, etc. n'intervienne pour influencer leur classement (Green, 1994).

Les travaux des psychologues ont mis en lumière de nombreuses sources de variation interindividuelles affectant l'apprentissage (Huteau, 1985). L'évaluation formative, puisqu'elle doit faciliter cet apprentissage en offrant aux sujets les meilleures conditions d'étude, devrait s'intéresser aussi bien à l'étude des styles cognitifs des sujets qu'aux facteurs de motivation qui leur sont propres, ou au mode d'apprentissage privilégié dans leur culture, etc. De nouvelles perspectives de recherche sur l'interaction aptitude-traitement ont été proposées par Rogosa (1991) et Snow (1991).

Le pôle de la Société

Beaucoup de travaux importants sur l'évaluation échapperaient à cette revue, si l'on n'ajoutait pas un pôle supplémentaire, celui de la sociologie (micro- ou macrosociologie), dont la vision d'ensemble permet de mieux comprendre les relations à l'évaluation de chacun des trois autres pôles: Apprenant, Enseignant et Savoir. Nous en citerons seulement quelques exemples à titre illustratif.

La sociologie de l'évaluation du système Apprenant

Les stratégies des parents. - L'approche macrosociologique des chercheurs de l'IREDU (Institut de recherche en économie de l'éducation) leur permet de dégager des lois relatives à l'évaluation scolaire, en se fondant sur une épistémologie objectiviste. C'est sur la base de relations statistiques que Duru-Bellat et Mingat (1987) démontrent entre autres que les parents de milieu social élevé sont prêts à prendre des risques pour conserver à leurs enfants des chances d'accès à l'enseignement supérieur. Au contraire, dans les milieux ouvriers, seul un pronostic brillant peut justifier qu'on entreprenne de longues études.

Des mises en corrélation du même type leur permettent d'étudier le fonctionnement des établissements plus ou moins sélectifs, en mettant en évidence les conséquences subtiles et souvent paradoxales de leurs politiques d'évaluation, compte tenu des stratégies différentes des parents de chaque milieu social.

Le métier d'élève. - L'approche microsociologique de Perrenoud (1984 et 1994) permet de mieux comprendre les buts, les valeurs, les enjeux et les stratégies des élèves, face à l'évaluation qu'ils subissent. Toutes sortes de façons se présentent pour eux de survivre dans ce milieu pauvre en renforcements positifs. La fuite, ou la politique du moindre effort, sont aussi «raisonnables» que le travail acharné, si l'on considère les satisfactions qu'un élève peut escompter à court terme. Seules des valeurs sociales héritées de son

milieu familial peuvent amener un élève à se dissocier du groupe classe qui cherche à négocier à la baisse les exigences à satisfaire.

On voit que, pour rendre compte du fonctionnement concret de l'évaluation scolaire, se référer aux seules situations d'examen standardisées serait bien insuffisant. L'explicitation des stratégies des acteurs est nécessaire pour comprendre la signification des relations statistiques qui apparaissent dans les analyses du flux d'ensemble des élèves.

La sociologie de l'évaluation du système Enseignant

L'évaluation de la lutte contre l'échec scolaire. - L'approche macrosociologique permet de mettre en évidence la résistance que le réel social oppose aux politiques volontaristes, aussi bien des enseignants que des autorités scolaires. Hutmacher (1993) a analysé en effet les résultats décevants des multiples efforts tentés à Genève depuis vingt ans pour réduire le taux de retard scolaire, dans un contexte social pourtant particulièrement favorable. Étude d'évaluation du «système enseignant», sa recherche est en même temps fondamentale pour comprendre le fonctionnement de l'évaluation des apprentissages à l'école.

Il apparaît bien que le taux de redoublement est relativement indépendant du niveau réel d'apprentissage. La réduction du nombre d'échecs est donc possible, si l'on intervient directement sur les démarches d'évaluation des enseignants, mais au prix d'autres tensions que le système scolaire ne parvient pas à supporter à long terme: la fréquence des redoublements retrouve vite son niveau initial.

Les indicateurs internationaux de l'enseignement (OCDE-CERI, 1991) s'inspirent du même paradigme objectiviste. Ils visent à contrôler l'effet de caractéristiques mesurables des systèmes scolaires sur les résultats de l'enseignement, pour faciliter le «pilotage» de la formation (Landsheere, 1994).

L'évaluation des projets d'établissement. - L'approche microsociologique, comme celle d'Huberman et Miles (1984) ou de Derouet (1987), permet de comprendre les multiples facteurs qui influencent le déroulement et les effets des programmes éducatifs d'ensemble, à partir de l'analyse de projets plus limités d'école ou de collège. L'examen comparatif par ces auteurs d'une série d'études de cas révèle notamment le rôle déterminant du directeur de l'établissement, dont l'attitude peut favoriser ou bloquer l'acceptation de l'innovation proposée, par les réactions qu'elle suscite dans le groupe des professeurs concernés.

La phase d'évaluation des résultats du projet est elle-même une aventure sociale, dont la conduite est à négocier entre les différents partenaires. La méthodologie de l'évaluation en général a beaucoup à apprendre de telles expériences, comme l'a montré Figari (1994): on ne peut espérer gérer convenablement l'introduction du changement dans une organisation sans tenir compte attentivement du point de vue des principaux intéressés.

La sociologie de l'évaluation du système Savant

L'évaluation des curriculums.- Jamais les responsables politiques n'ont autant discuté avec les chercheurs en sciences sociales des finalités et des objectifs de l'éducation. Comment préparer les jeunes à la société du prochain millénaire? Comment déterminer les apprentissages essentiels? Comment formuler les plans d'études? Quel poids accorder à chaque discipline?

Malgré certains efforts de systématisation (Darling-Hammond et Snyder, 1992; Perret et Perrenoud, 1990), l'évaluation des curriculums reste une démarche tâtonnante, où les approches descriptives et prescriptives s'opposent nécessairement. La pratique de la Suisse romande montre que la combinaison de phases d'observation objective et de consultations d'experts permet au moins de négocier une évolution raisonnée des contenus d'enseignement (par exemple, Weiss, 1992).

L'évaluation des savoirs.- Ces changements globaux de plans d'études révèlent que certains domaines de savoirs deviennent prioritaires par rapport à d'autres connaissances, jugées moins essentielles. Des facteurs comme le progrès technique et la spécialisation professionnelle expliquent ce déplacement des savoirs, d'après Girod (1991). Des enquêtes détaillées sur la diffusion des résultats de recherche, comme celle de Huberman et Gather Thurler (1991), éclairent aussi cette transformation des hiérarchies d'importance. L'étude de la pondération des disciplines dans le curriculum de chaque pays permettrait sans doute d'expliquer certains des résultats obtenus dans les comparaisons internationales.

En bref, ces quelques exemples de recherches sociologiques tendent à prouver que l'on ne peut comprendre et faire progresser l'évaluation en éducation sans avoir une vision systémique et compréhensive des problèmes. Ceci nécessite de prendre en compte l'influence des réactions des acteurs sociaux concernés dans les décisions envisagées.

Quelles lignes de force finalement ?

Si l'on passe en revue les travaux mentionnés dans le tour d'horizon précédent, il nous semble qu'une constante se dégage assez nettement pour la majorité d'entre eux: la prise en compte du point de vue subjectif des intéressés, par opposition à une approche purement comportementaliste. Ce point de vue ressort à tous les niveaux :

1. *Pour l'enfant ou l'adulte apprenant*, on essaye de partir de ses représentations du domaine étudié et de tenir compte de ses projets et stratégies d'études, en l'aidant à découvrir les objectifs à poursuivre dans son cas particulier, à s'auto-évaluer, et à négocier des contrats d'apprentissage qui développent son autonomie.
2. *Pour les parents*, on accepte mieux leurs stratégies propres pour l'orientation des études de leurs enfants. Au lieu de vouloir prioritairement assurer une égalité des chances formelle, on cherche à promouvoir le plus grand nombre d'enfants possible, en repoussant à plus tard dans le cours des études les décisions de sélection à portée professionnelle.
3. *Pour les enseignants*, on reconnaît qu'ils ont besoin d'être appuyés dans leur désir d'individualiser enseignement et évaluation. On essaye de créer pour eux des moyens d'enseignement appropriés aux différents styles cognitifs et des moyens d'évaluation à visée diagnostique. Ces deux types d'instruments sont conçus en accord avec les étapes de l'apprentissage de chaque discipline et les obstacles épistémologiques correspondants.
4. *Pour les dirigeants*, enfin, on admet qu'ils soient moins intéressés par une évaluation purement objective que par une information qui peut les aider à gérer la part du système d'enseignement dont ils ont la charge. Ils ont particulièrement besoin aujourd'hui de méthodes d'évaluation pour réguler le fonctionnement des établissements (Lecointe et Rebinguet, 1990) et pour adapter le contenu des curriculums (Lewy, 1991). Ces méthodes font nécessairement appel à la négociation de référentiels avec les autres partenaires sociaux.

Le choix de ces quelques lignes de force est naturellement subjectif et ne signifie nullement que les autres directions de recherche et de développement soient moins intéressantes ou moins utiles. Ces dernières nous semblent simplement moins cohérentes entre elles, plus dispersées dans leurs appro-

ches, que les études qu'on vient de relever, dont le maître mot nous semble être «la compréhension» du point de vue de chaque acteur.

Les progrès à accomplir

Autres secteurs d'activité

La revue précédente s'est surtout centrée sur les progrès réalisés dans le domaine de l'évaluation des apprentissages en milieu scolaire. Ce choix ne doit pas faire oublier que les pratiques d'évaluation, dans d'autres secteurs d'activité, ont également besoin d'être approfondies et améliorées. L'évaluation des enseignants, de l'enseignement, des institutions, des programmes d'étude et des didactiques, du matériel scolaire et des didacticiels, etc., mérite également une grande attention. Ces évaluations reposent sur des régulations complexes, impliquant non seulement des rapports au savoir et à la technologie, mais aussi des rapports humains, sociaux, institutionnels et politiques. Les mécanismes régissant ces évaluations doivent être clairement établis et étudiés, afin d'assurer plus de rigueur, plus de transparence et plus d'équité dans les processus décisionnels.

Interactions entre secteurs

De plus en plus, nous réalisons que l'évaluation des apprentissages, si nous souhaitons la comprendre et l'utiliser à des fins qui permettent d'améliorer les systèmes d'éducation, doit s'accompagner d'évaluations régulières aux autres niveaux du système éducatif. Un système d'examen nationaux uniformes, par exemple, n'est guère utile s'il ne s'accompagne pas d'une politique d'évaluation institutionnelle permettant d'assurer un suivi approprié et de choisir les correctifs à apporter pour améliorer les résultats d'apprentissage. Ces correctifs peuvent porter sur les moyens d'enseignement, la formation des enseignants, etc. En multipliant les prises d'information et les liens entre celles-ci, l'évaluation en éducation deviendra plus systémique et ainsi plus efficace, même si aucune technique ne saurait éliminer complètement les erreurs d'évaluation et de prise de décision.

Évaluation de l'évaluation

Logiquement, la démarche de contrôle régulateur prônée pour l'évaluation devrait pouvoir s'appliquer à elle-même, pour qu'elle puisse se perfectionner. L'implantation du bulletin critérié au Québec dans les années 80 est certainement un exemple d'une situation où l'évaluation continue d'un processus de changement dans la transmission des résultats aurait permis d'éviter d'avoir parfois à faire marche arrière (Conseil supérieur de l'éducation,

1992). De meilleures modalités de transmission des résultats auraient pu être imaginées. Surtout, de meilleures procédures auraient dû être suivies pour introduire un tel changement auprès des directions d'école, des enseignants et des parents.

La Direction de l'évaluation et de la prospective, en France, interroge régulièrement les enseignants sur les différents dispositifs d'aide à l'évaluation mis à leur disposition (MEN-DEP, 1996). L'opinion des usagers est majoritairement favorable.

L'exemple anglo-saxon

Pour proposer des améliorations, cependant, il ne suffit pas de constater les dysfonctionnements de façon précise. Il faut imaginer des solutions inédites. C'est pourquoi il faut continuer à effectuer des recherches dans le domaine de l'évaluation, à l'exemple des pays anglo-saxons.

Aux États-Unis en particulier, l'évaluation est une industrie impliquant de grandes maisons d'éditions (*McGraw-Hill, Laurence Erlbaum, Psychological Corporation*), et des organismes de recherche privés ou semi-publics (*American College Testing, Educational Testing Service*). Des sommes considérables sont investies pour développer de nouveaux produits, ou pour satisfaire des critères élevés sur le plan éthique. En effet, cette industrie est contrôlée par de nombreuses lois ou codes professionnels de déontologie, et elle est soumise à une abondante jurisprudence. Enfin, de nombreuses revues savantes nord-américaines traitent de questions d'évaluation, en particulier *Educational Measurement: Issues and Practice, Educational Researcher, Journal of Educational Measurement, Journal of Applied Psychological and Educational Measurement*, pour ne nommer que celles qui se préoccupent principalement de l'évaluation des apprentissages.

Pour répondre aux nouvelles exigences, des efforts conjoints réunissant l'*American Educational Research Association*, l'*American Psychological Association* et le *National Council for Measurement in Education* ont permis de développer une série de directives en ce qui concerne les différents aspects de la préparation, de la correction, de la notation et de la validation des instruments de mesure psychologique ou éducationnelle. Ce sont les «*Standards for educational and psychological testing*» (*American Psychological Association*, 1985). Au Canada, un ouvrage plus général intitulé *Principes d'équité relatifs aux pratiques d'évaluation des apprentissages scolaires au Canada* a également été produit en 1993 par le *Center for Research in Applied Measurement and Evaluation*.

En acceptant que «tout se complique... »

En jetant ce regard en arrière sur la décennie écoulée, nous ne pouvons éviter de comparer l'idée que nous nous faisons alors de l'évaluation à celle que nous en avons aujourd'hui, et notre point de vue ancien nous semble étrangement naïf.

Notre état de la question a servi à présenter tous les développements nouveaux qu'a connus le champ des pratiques. Des enseignants, des établissements, des systèmes scolaires entiers ont opté pour de nouvelles façons de faire leurs évaluations, et ils se heurtent naturellement à des obstacles imprévus, générateurs d'innovations supplémentaires. On assiste à la multiplication des champs d'application, comme des fonctions attendues de l'évaluation, et il en résulte naturellement une multiplication des modèles proposés.

Nous avons évoqué le triomphe de l'approche cognitiviste/néo-constructiviste et son impact sur les autres sciences sociales, où le réel étudié devient multiple, réfracté au travers d'une multitude de consciences individuelles. Les scientifiques sont obligés de réintroduire la subjectivité dans leur modèle du monde, et donc de faire une place aux philosophes et aux questions éthiques.

«Rien n'est simple» désormais, «tout se complique», comme disait Sempé. Pourtant, nous ne pouvons douter que ce soit la marque d'un progrès, parce que notre vision du réel s'enrichit et se diversifie. Nous commençons à penser la complexité.

RÉFÉRENCES

- Allal, L. (1991). Vers une pratique de l'évaluation formative. Bruxelles: De Boeck.
- Allal, L. (1979). Stratégies d'évaluation formative : conceptions psychopédagogiques et modalités d'application. In L. Allal, J. Cardinet & Ph. Perrenoud (éds), L'évaluation formative dans un enseignement différencié. Berne : Peter Lang.
- Allal, L., Bain, D. & Perrenoud, Ph. (1993). Évaluation formative et didactique du français. Neuchâtel : Delachaux et Niestlé.
- American Psychological Association (1985). Standards for educational and psychological testing. Washington, DC : American Psychological Association.
- Ardoino, J. & Berger, G. (1989). D'une évaluation en miettes à une évaluation en actes : le cas des universités. Paris : Matrice et Andsha.

- Astolfi, J.-P. (1992). L'école pour apprendre. Paris : ESF.
- Astolfi, J.-P. & Develay, M. (1989). La didactique des sciences. Paris : PUF.
- Audigier, F. (dir.) (1987). Actes de la II^e Rencontre nationale sur la didactique de l'histoire et de la géographie. Paris : INRP.
- Bain, D. & Schneuwly, B. (1993). Pour une évaluation formative intégrée dans la pédagogie du français : de la nécessité et de l'utilité de modèles de référence. In L. Allal, D. Bain & Ph. Perenoud (dir.), Évaluation formative et didactique du français (pp. 51-79). Neuchâtel : Delachaux et Niestlé.
- Baker, E.L., O'Neil, H.F. & Linn, R.L. (1993). Policy and validity prospects for performance-based assessment. American Psychologist, 48(12), 1210-1218.
- Bandler, R. & Grinder, J. (1990). Un cerveau pour changer : la programmation neurolinguistique. Paris : InterÉditions.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. Educational Psychologist, 28(2), 117-148.
- Barkatoolah, A. (1987). Expériences nord-américaines en reconnaissance des acquis (Québec, États-Unis). Pratiques de formation. Numéro spécial.
- Barth, B.-M. (1993). Le savoir en construction : Former à une pédagogie de la compréhension. Paris : Retz.
- Bednarz, N. & Garnier, C. (dirs) (1989). Construction des savoirs. Obstacles et conflits. Ottawa : Cirade-Agence d'Arc Inc.
- Bertrand, O. (éd.) (1996). Qualifications et compétences professionnelles dans l'enseignement technique et la formation professionnelle : évaluation et certification / sous la dir. de Marianne Durand-Drouhin. Paris : OCDE.
- Bertrand, R. & Jeanrie, Ch. (1994). Le rôle des théories modernes dans la résolution des problèmes de mesure. Mesure et évaluation en éducation, 17(2), 1-19.
- Black, H. & Deakins, G. (1990). Evaluation from both sides of the same fence. In : Scottish Council for Research in Education, Council of Europe, The evaluation of educational programmes : methods, uses and benefits (pp. 11-23). Amsterdam : Swets and Zeitlinger.
- Bonniol, J.-J. (1981). Influence de l'explicitation des critères utilisés sur le fonctionnement des mécanismes d'évaluation d'une production scolaire. Bulletin de psychologie, 35(353), 173-186.
- Bonniol, J.-J. (1986). Recherche et formations : pour une problématique de l'évaluation formative. In J.-M. De Ketele (éd.), L'évaluation : approche descriptive ou prescriptive?. Bruxelles : De Boeck.
- Bronckart, J.-P. (1985). Le fonctionnement des discours. Un modèle psychologique et une méthode d'analyse. Neuchâtel : Delachaux et Niestlé.
- Brousseau, G. (1986). Fondements et méthodes de la didactique des mathématiques. Recherches en didactique des mathématiques, 7(2), 33-115.

- Cardinet, J. (1989). Évaluer sans juger. Revue française de pédagogie, 88, 41-52.
- Cardinet, J. (1994). Une typologie des modèles d'évaluation institutionnelle. In D. Bain (coord.), Fonctionnement de l'évaluation dans la formation - Évaluation du fonctionnement de la formation (pp. 73-78). Actes du Colloque de l'ADMEE 1994. Genève : CRPP.
- Carraher, T.N., Carraher, T.W. & Schliemann, A.D. (1985). Mathematics in street and school. British Journal of Developmental Psychology, (3), 21-29.
- CCEMPC (1993). Programme intégré : plans de référence pour un projet pédagogique d'école fondamentale. Liège : Conseil central de l'enseignement maternel et primaire catholique.
- Centre for Research in Applied Measurement and Evaluation (1993). Principes d'équité relatifs aux pratiques d'évaluation des apprentissages scolaires au Canada. Edmonton : University of Alberta.
- Chevallard, Y. (1985). La transposition didactique. Grenoble : La Pensée Sauvage.
- Conseil supérieur de l'éducation (1992). Évaluer les apprentissages au primaire : un équilibre à trouver. Avis au ministre de l'Éducation, Québec.
- Cronbach, L., Linn, R., Brennan, R. & Haertel, E. (1995). Generalizability analysis for educational assessments. Evaluation Comment, Summer issue, 1-29. (Available electronically from <http://www.cse.ucla.edu>.)
- Darling-Hammond, L. & Snyder, J. (1992). Curriculum studies and the traditions of inquiry : the scientific tradition. In Ph. W. Jackson, Handbook of research on curriculum (pp. 41-78). New-York : Macmillan.
- Darmuzey, Ch. & A. (1988). Un outil de régulation. Animation et éducation, 84, 13-16.
- Dechef, H. & Laveault, D. (1993). Étude du fonctionnement différentiel des items à l'aide des méthodes du khi-carré, de Mantel-Haenszel et Logit. Mesure et évaluation en éducation, 16 (1-2), 5-28.
- De Ketele, J.-M. (1993). L'évaluation conjugée en paradigmes. Revue française de pédagogie, 103, 59-80.
- Derouet, J.-L. (1987). Une sociologie des établissements scolaires : les difficultés de construction d'un nouvel objet scientifique. Revue française de pédagogie, 78, 86-108.
- Doyon, C. (1991). Faire participer l'élève à l'évaluation de ses apprentissages. Laval, Québec : Beauchemin.
- Duru-Bellat, M. & Mingat, A. (1987). Le redoublement au Collège, en France : éléments pour une évaluation. Dijon : IREDU.
- Eugène, J.-P., Loncle, J.-C. & Montméas, L. (1987). Brevet de technicien agricole-Modalités d'opérationnalisation et d'évaluation des modules de qualification.

Documents INRAP N° 65. Dijon : Institut national de recherches et d'applications pédagogiques du ministère de l'Agriculture.

Figari, G. (1994). Évaluer : quel référentiel?. Bruxelles : De Boeck.

Fournier, C. & Laveaut, D. (1994). Liens entre anticipation, auto-évaluation et le résultat à un examen de rendement scolaire. Revue des sciences de l'éducation, 20(3), 427-442.

Frechtling, J.A. (1991). Performance assessment : moonstruck or the real thing? Educational Measurement : Issues and Practice, 10(4), 23-25.

Frederiksen, J.R. & Collins, A. (1989). A systems approach to educational testing. Educational Researcher, 18(9), 27-32.

Giordan, A. (dir.) (1989). Psychologie génétique et didactique des sciences. Berne : Peter Lang.

Giordan, A. (1995). Les nouveaux modèles pour apprendre : dépasser le constructivisme? Perspectives, 35(1), 109-127.

Giordan, A. & De Vecchi, G. (1987). Les origines du savoir. Des conceptions des élèves aux concepts scientifiques. Neuchâtel : Delachaux et Niestlé.

Girod, R. (1991). Le savoir réel de l'homme moderne : essais introductifs. Paris : PUF.

Green, B.F. (1994). Differential item functioning : techniques, findings and prospects. In D.Laveaut, B.D. Zumbo, M.E. Gessaroli & M.W. Boss (éds), Modern Theories of Measurement : Problems and Issues (pp. 141-162). Ottawa, Canada: Edumetrics Research Group, University of Ottawa.

Grisay, A. (1988). Du mythe de la «bonne école» à la réalité (fuyante) de «l'école efficace», Liège : Service de pédagogie expérimentale de l'Université de Liège.

Grossen, M. (1988). L'intersubjectivité en situation de test. Cousset (Fribourg) : Éditions Delval.

Groupe EVA (1991). Évaluer les écrits à l'école primaire. Paris : Hachette.

Guilbert, M., Hibon, M., Le Lièvre-Bourdoin, L., Ouanas, H. & Tavernier, R. (1995). L'évaluation à l'école élémentaire : Guide du maître. Paris : Bordas.

Hadji, C. (1992). L'évaluation des actions éducatives. Paris : Presses Universitaires de France.

Hambleton, R.K. & Swaminathan, H. (1985). Item response theory : principles and applications. Boston : Kluwer.

Hambleton, R.K., Swaminathan, H. & Rogers, J. (1991). Fundamentals of item response theory. Newbury Park : Sage.

Huberman, M. & Gather Thurler, M. (1991). Diffuser les savoirs. Eléments de base et modes d'emploi pour chercheurs et praticiens. Berne : Lang.

Huberman, M. & Miles, M. (1984). Innovation up close : how school improvement works. New York : Plenum Press.

- Huteau, M. (1985). Les conceptions cognitives de la personnalité. Paris : PUF.
- Hutmacher, W. (1993). Quand la réalité résiste à la lutte contre l'échec scolaire. Genève : Service de la recherche sociologique.
- Jaeger, R.M. (1994). The psychometric demands of testing for licensure and certification. In D.Laveault, B.D. Zumbo, M.E. Gessaroli & M.W. Boss (éds), Modern Theories of Measurement : Problems and Issues (pp. 141-162). Ottawa, Canada: Edumetrics Research Group, University of Ottawa.
- Johnson, S. (1989). Évaluation de la comparabilité des notations entre jurys d'examens. Mesure et évaluation en éducation, 12(1), 5-22.
- Jonnaert, P. & Laveault, D. (1994). Évaluation de la familiarité de la tâche : quelle confiance accorder à la perception de l'élève. Revue des sciences de l'éducation, 20(2), 271-291.
- La Garanderie, A. (de) (1980). Les profils pédagogiques. Paris : Le Centurion.
- Landsheere, G. (de) (1994). Le pilotage des systèmes de formation. Bruxelles : De Boeck.
- Laveault, D. & Fournier, C. (1990). Évaluation par objectifs : une approche métacognitive. Mesure et évaluation en éducation, 13(1), 57-74.
- Laveault, D. (Éd.) (1992). Les pratiques d'évaluation en éducation. Montréal : Éditions de l'ADMEE, 49-61.
- Lecoite, M. & Rebinguet, M. (1990). L'audit de l'établissement scolaire. Paris : Les Éditions d'Organisation.
- Lejong, M., Detheux, M. & Bal, R. de (1997). Épreuves d'évaluation sur les compétences de base en mathématique : 4^e année primaire : recherche en vue de l'amélioration des pratiques d'évaluation-bilan et d'une réduction des redoublements. Liège : Université, Laboratoire de pédagogie expérimentale.
- Lewy, A. (éd.) (1991). The international encyclopedia of curriculum. Oxford : Pergamon Press.
- Linn, R.L. & Burton, E. (1994). Performance-based assessment : implications of task specificity. Educational Measurement : Issues and Practice, 13(1), 5-8.
- Lussier-Charette, F. (1987). L'intégration des objectifs des programmes d'études : une expérience stimulante en troisième année du primaire. Vie pédagogique, 51, 21-23.
- McNicoll, A. (1996). Étudgen : logiciel de généralisabilité. Québec : CESSUL, Université Laval.
- Mehrens, W.A. (1992). Using performance assessment for accountability purposes. Educational Measurement : Issues and Practice, 11(1), 3-9.
- MEN-DEP (1996). L'usage des dispositifs et outils d'aide à l'évaluation dans l'enseignement primaire et secondaire. Note d'information, 96.46.

- Messick, S. (1989). Validity. In R.L. Linn (éd.), Educational measurement (3rd edition) (pp. 13-95). New York : Macmillan.
- Meyer, N. & Berger, G. (1988) Reconnaissance des acquis : étude à caractère documentaire. (2^e éd.). Paris : Université Paris VIII, Formation permanente.
- Ministère de l'Éducation du Québec (1993). Des collèges pour le Québec du XXI^e siècle. Québec : Direction générale de l'enseignement collégial.
- Mislevy, R.J. & Bock, R.D. (1986). BILOG : Maximum likelihood item analysis and test scoring with logistic models. Mooresville, IN : Scientific Software Inc.
- Montméas, L., Dauvisis, M.-C., Chautard, P., Courtas, J. & Frossard, G. (1989). Guide méthodologique pour gérer le contrôle en cours de formation. Documents INRAP DPE 089. Dijon : Institut national de recherches et d'applications pédagogiques du ministère de l'Agriculture.
- Morissette, D. & Gingras, M. (1989). Évaluer des attitudes : planifier, intervenir, évaluer. Bruxelles/Sainte-Foy : DeBoeck/PUL.
- Morissette, D. (1993). Les examens de rendement scolaire. Sainte-Foy : Presses de l'Université Laval.
- Nadeau, M.-A. (1988). L'évaluation de programme : théorie et pratique. Sainte-Foy : Presses de l'Université Laval.
- Noël, B. (1991). La métacognition. Bruxelles : De Boeck.
- Nunziati, G. (1989). Formation de formateurs en évaluation sommative et formative: rapport intermédiaire. Aix-en-Provence : Unité d'enseignement et de recherche.
- Nunziati, G. (1990). Pour construire un dispositif d'évaluation formatrice. Cahiers pédagogiques, 280, 47-64.
- OCDE-CERI (1991). Manuel sur les indicateurs internationaux de l'enseignement. Coopération internationale en éducation, 1992(numéro spécial S3/92), 5-49.
- Paris, S. G. & Ayres, L. R. (1994). Becoming reflective students and teachers with portfolios and authentic assessment. Washington : APA Books.
- Paulson, L. (1996). Portfolio de mathématiques à l'école primaire : mode d'emploi. Adapt. française par S. Foster et Ch. Sandoz. Neuchâtel : IRDP.
- Perrenoud, Ph. (1984). La fabrication de l'excellence scolaire. Genève : Droz.
- Perrenoud, Ph. (1994). Métier d'élève et sens du travail scolaire. Paris : ESF.
- Perret, J.-F. & Perrenoud, Ph. (éds) (1990). Qui définit le curriculum, pour qui?. Cousset (Fribourg) : IRDP-Delval.
- Perret-Clermont, A.-N. & Nicolet, M. (éds) (1988). Interagir et connaître. Enjeux et régulations sociales dans le développement cognitif. Cousset (Fribourg) : Delval.
- Popham, W.J. (1993). Educational testing in America : what's right, what's wrong?. Educational Measurement : Issues and Practice, 12(1), 11-14.

- Reckase, M.D. (1995). Portfolio assessment : a theoretical estimate of score reliability. Educational Measurement : Issues and Practice, 14(1), 12-14.
- Resnick, L.B. & Ford, W.W. (1981). The psychology of mathematics for instruction. Hillsdale, NJ : Erlbaum.
- Rogosa, D. (1991). A longitudinal approach to ATI research : models for individual growth and models for individual differences in response to intervention. In R. E. Snow & D. E. Wiley (éds), Improving inquiry in social science. A volume in honor of Lee J. Cronbach (pp. 221-248). Hillsdale, NJ : Lawrence Erlbaum.
- Romainville, M. (1993). Savoir parler de ses méthodes. Bruxelles : De Boeck.
- Scallon, G. (1988). L'évaluation formative des apprentissages (t. 2 : l'instrumentation). Sainte-Foy : Presses de l'Université Laval.
- Scallon, G. (1992). L'impact des pratiques d'évaluation sur la motivation des élèves: le poulx de la recherche. Monographies en mesure et évaluation. Sainte-Foy : Université Laval.
- Schubauer-Leoni, M.-L., Bell, N., Grossen, M. & Perret-Clermont, A.-N. (1989). Problems in the assessment of learning : the social construction of questions and answers in the scholastic context. International Journal of Educational Research, 13(6), 671-684.
- Sevikian, M. (1986). Appropriation des critères et construction cognitive : première approche chez des enfants d'école maternelle. Mémoire de maîtrise, Université de Provence, Aix.
- Snow, R. (1991). The concept of aptitude. In R. E. Snow & D. E. Wiley (éds), Improving inquiry in social science. A volume in honor of Lee J. Cronbach (pp.249-284). Hillsdale, NJ : Lawrence Erlbaum.
- Tochon, F. (1993). L'enseignante experte, l'enseignant expert. Paris : Nathan.
- Vermersch, P. (1994). L'entretien d'explicitation. Paris : ESF.
- Veslin, O. & Veslin, J. (1992). Corriger des copies : évaluer pour former. Paris : Hachette.
- Vial, M. (1987). Statut de la carte d'étude dans un dispositif d'évaluation-régulation. Pratiques, 53, 59-73.
- Viennot, L. (1979). Le raisonnement spontané en dynamique élémentaire. Paris : Herman.
- Walberg, H. J. & Haertel, G. D. (1990). The international encyclopedia of educational evaluation. Oxford : Pergamon Press.
- Weiss, J. (éd.) (1991). L'évaluation : problème de communication. Neuchâtel : Delval / IRDP.

- Weiss, J. (éd.) (1992). Français 91. Colloque organisé par la Conférence intercantonale des chefs des Départements de l'instruction publique de la Suisse romande et du Tessin : rapport final. Neuchâtel : IRDP.
- Weiss, J. (éd.) (1996). Vers une conception cohérente de l'évaluation pour la scolarité obligatoire en Suisse romande et au Tessin. Neuchâtel : IRDP.
- Wirthner, M., Martin, D. & Perrenoud, Ph. (dir.) (1991). Parole étouffée, parole libérée : Fondements et limites d'une pédagogie de l'oral. Neuchâtel : Delachaux et Niestlé.
- Wolfs, J.-L. (1993). Analyse des questions d'évaluation anticipées par les étudiants: approche exploratoire. Mesure et évaluation en éducation, 15(4), 31-51.
- Ysewijn, P. (1996). GT : Logiciel pour études de généralisabilité. Bercher (CH-1038): Pierre Ysewijn.