

## L'évaluation conjuguée en paradigmes

*Jean-Marie De Ketele*

Dans les lignes qui suivent, le lecteur ne doit pas espérer trouver un résumé des travaux sur l'évaluation, mais bien une synthèse. Il s'agit donc d'une sélection et d'une mise ensemble d'éléments à partir d'un regard personnel porté sur un nombre relativement important de travaux menés dans le domaine de l'évaluation. Nous n'avons pas voulu mentionner tous les travaux auxquels nous avons eu accès. Bien au contraire, nous avons préféré écrire cet article en prenant une certaine hauteur, le plus souvent de mémoire, en ne consultant les documents qu'après coup à des fins de vérification ou de précision des références.

Pour articuler des éléments aussi nombreux que ceux que l'on rencontre dans les travaux sur l'évaluation, il faut choisir un point d'entrée parmi de nombreux points d'entrée possible, tout aussi pertinents que celui que nous avons choisi : tenter de dégager les paradigmes qui ont guidé une palette assez large d'auteurs. Selon Kuhn (cité par Doyle, 1986), « un paradigme est un cadre implicite qui définit des problèmes, des méthodes et des solutions légitimes pour une communauté scientifique ». En essayant de dégager les paradigmes des travaux sur l'évaluation, nous avons conscience que nous allons opérer quelques réductionnismes : bien des auteurs ont vu leur pensée évoluer au fil du temps ; parfois même, certains auteurs ont utilisé des paradigmes différents en fonction des objets qu'ils avaient à évaluer ou selon les contraintes du contexte évaluatif. Sur ce dernier aspect, il suffirait de confronter les écrits et les pratiques réelles des experts de l'évaluation lorsqu'ils sont amenés à évaluer leurs propres étudiants ! Que l'on me pardonne donc le réductionnisme de ces lignes : il n'a d'autre prétention que d'articuler clairement quelques idées de force.

Ayant passé en revue quelques paradigmes importants à nos yeux, nous tenterons d'examiner dans quelle mesure ils peuvent se compléter — bien plus que s'opposer — dans une vision élargie de l'évaluation et quelles sont les questions qui se posent, espérant ainsi ouvrir de nouveaux horizons ou faire découvrir des horizons encore peu explorés.

## QUELQUES GRANDS PARADIGMES DE L'ÉVALUATION

### Le paradigme de l'intuition pragmatique

Ce paradigme est à la base de la majorité des pratiques de l'évaluation, même encore actuellement malgré le nombre important d'écrits qui les ont dénoncées et malgré le nombre important de formations qui ont tenté de les remplacer par d'autres pratiques.

L'acte évaluatif est ici un acte intuitif à bien des titres. Pas besoin de définir avec précision, ni ses objectifs, ni l'univers de référence des situations possibles de l'évaluation, ni les critères à prendre en considération, ni la façon de synthétiser les informations disponibles, ni la façon de les interpréter, ni la façon d'utiliser les résultats. L'évaluation est un acte syncrétique étroitement lié à la personne de l'évaluateur et admis comme tel, dans la mesure où l'évaluateur est celui qui a mené l'apprentissage, revendique être la personne la mieux placée pour bien connaître les performances de ses élèves ou de ses étudiants, prend ses responsabilités avec conscience professionnelle.

C'est aussi un acte pragmatique fortement déterminé par les exigences du contexte. Puisque les parents demandent à être informés sur les performances de leurs enfants et qu'ils aiment voir celles-ci situées par rapport aux autres enfants de la classe ou du moins par rapport à un système de référence familial qui leur permet de supputer les chances de réussite de l'année, les enseignants utilisent un système de notes qui les dispensent de justifier les contenus précis des épreuves et qui offrent certaines apparences d'objectivité.

Puisque les contraintes de temps sont là et qu'il vaut mieux consacrer davantage de temps à l'apprentissage, les enseignants n'éprouvent guère le besoin de consacrer beaucoup de temps à la préparation des situations d'évaluation. Ayant constaté que les notes constituaient leur principale arme pour motiver leurs élèves, voire, dans certains cas, pour faire respecter la discipline, les enseignants trouvent cette forme d'évaluation bien pratique et se demandent par quoi ils pourraient bien la remplacer si on la leur enlevait.

Le visage des pratiques explicites de ce paradigme est bien celui de l'évaluation sommative, puisque la note somme un ensemble d'informations récoltées occasionnellement et intuitivement. On pourrait en conclure assez vite que la fonction première est de contribuer à la certification de la réussite des élèves. Il nous semble que c'est là une inférence trop rapide. Comme les travaux de Perrenoud (1979, 1984, 1991) l'ont bien montré, de telles pratiques ont pour fonction première la gestion sociale : gestion de la classe par le pouvoir de motivation et de sanction de la note ; gestion institutionnelle car l'enseignant peut être jugé par ses collègues ou par les administrateurs à partir des profils de notes qui seront publiques ; gestion des relations avec les parents que l'on désire tenir à l'écart de certaines formes de pression (Montandon, 1991). Dans ce cadre, la fonction de certification n'est qu'une des fonctions secondes, subordonnée à la fonction de gestion sociale.

S'il est vrai que de nombreux travaux sur l'évaluation dénoncent explicitement de telles pratiques ou tendent à vouloir les remplacer ou les faire évoluer en proposant d'autres paradigmes, on trouve cependant des arguments dans la littérature spécialisée pour défendre un tel paradigme. Citons deux lignées de travaux.

La première rassemble des recherches qui se sont penchées sur les prédicteurs de la réussite. La plupart des études montrent que les notes scolaires antérieures sont de meilleurs prédicteurs que les résultats à des tests objectifs, que ce soient des tests d'aptitude, de personnalité ou de performances scolaires standardisés.

La seconde lignée de travaux a trouvé une de ses meilleures expressions dans la communication de Weiss au colloque de l'ADMEE à Bruxelles en 1985 sous le titre évocateur « La subjectivité blanchie ? » (publié in De Ketele, 1986). L'auteur se pose la question de savoir si finalement l'approche intuitive et pragmatique des enseignants ne vaut pas celle de l'approche des experts de l'évaluation qui ont trop tendance à dissocier formation et évaluation, à dissocier évaluation formative et certificative — dissociation idéalement possible mais socialement improbable —, à promouvoir des conceptions qui n'en sont pas moins exemptes de biais (arbitraire dans le choix des objectifs, des critères, ...). Finalement, le paradigme sous-jacent aux pratiques de la plupart des enseignants n'en vaut-il pas bien d'autres, car il est celui du réalisme et de l'utilité sociale : les évaluations des enseignants sont des appréciations homogènes et concordantes parce que socialement déterminées ; elles distinguent les deux types d'élèves qui réussissent (l'élève actif, sociable et intelligent ; l'élève appliqué et discipliné) et les deux types d'élèves qui échouent (l'élève passif, replié sur lui-même et peu doué ; l'élève peu travailleur, dissipé et indiscipliné), comme l'ont montré de nombreux travaux de recherches (Mollo, 1970 ; Kaufman, 1976 ; Londeix, 1982 ; Gilly, 1980) ; elles pronostiquent les réussites scolaires ultérieures à court et à long terme ; elles sont modulées par des pratiques correctrices lorsque les comportements observés ne sont pas conformes aux résultats attendus, ce qui explique leur bonne validité pronostique...

### **Le paradigme docimologique**

Le paradigme docimologique est essentiellement orienté vers par la fidélité ou la fiabilité des évaluations. Au départ, il a été surtout le fait de personnes formées à une certaine forme de pensée scientifique issue des travaux de Claude Bernard, des psychophysiciens, des travaux en psychologie expérimentale de l'apprentissage et de la pédagogie expérimentale. Piéron (1963) a été sans doute la figure la plus illustre de la première phase du mouvement docimologique, phase essentiellement critique qui a consisté à mettre en évidence les principaux biais présents dans les « examens » (d'où l'origine du terme de docimologie, science des examens). Dans une deuxième phase, on a assisté à une tentative d'explication de la production des résultats de l'évaluation et de ces biais. Les travaux de l'Ecole d'Aix-en-Provence autour de Bonniol, Noizet et Caverni (Bonniol, 1972 ; Bonniol, Caverni et Noizet, 1972a et 1972b ; Noizet et Caverni, 1978) contestent la réduction de l'évaluation à sa dimension sommative, l'assimilation de l'évaluation aux examens, la confusion des objectifs de l'évaluation et des objectifs pédagogiques. Pour eux, l'évaluation est une activité de comparaison entre une production scolaire à évaluer et un modèle de référence, comparaison qui est influencée par des déterminants systématiques, qui, tantôt se réfèrent à des caractéristiques scolaires, tantôt à des caractéristiques de personnalité, tantôt à des caractéristiques sociales. Dans une troisième phase, on assiste à des phénomènes contradictoires. D'une part, les travaux docimologiques tendent à diminuer et la docimologie est remise en question par de nombreux auteurs (voir par exemple Dauvisis, 1988). D'autre part, on assiste à un

mouvement tendant à créer une éduométrie, c'est-à-dire une science de la mesure dans le champ des sciences de l'éducation. Ce mouvement trouve son origine dans une double contestation : la mesure dans le champ de l'éducation ne peut se résoudre par les seuls modèles psychométriques ; le concept de fidélité est trop exclusivement pensé en termes de différenciation des personnes. Lancée et développée par des auteurs bien connus par ailleurs pour leur contribution à l'évaluation formative (Cardinet, Tourneur, Allal, 1976 ; Scallon, 1981), la théorie de la généralisabilité est représentative de ce mouvement. Il s'agit d'un cadre plus englobant qui distingue quatre types de différenciation : la différenciation des élèves ou la mesure de leurs traits distinctifs (les modèles les plus classiques sont de ce type) ; la différenciation des objectifs et des domaines d'enseignement ; la différenciation des conditions d'apprentissage et des facteurs d'enseignement ; la différenciation des niveaux successifs d'un apprentissage. Selon le type de différenciation recherché, les plans d'observation et les estimations de fidélité et de marges d'erreurs sont différents. On trouvera une synthèse de cette théorie dans « Assurer la mesure » (Cardinet et Tourneur, 1985).

Même si la docimologie a vu son champ s'élargir et si les auteurs les plus récents n'assimilent pas mesure et évaluation, il n'en reste pas moins que les travaux portent presque tous sur la « note » ou la « mesure ». D'autres paradigmes vont naître et coexister parallèlement au paradigme docimologique.

### **Le paradigme sociologique**

Ce paradigme s'inscrit dans la lignée des travaux de sociologues qui se sont intéressés au rôle que l'école jouait dans les mécanismes de reproduction sociale. Ainsi, Bourdieu et Passeron (1964, 1970) considèrent l'école comme le moyen de légitimer les inégalités culturelles ; Baudelot et Establet (1971, 1975) voient l'école comme un appareil qui permet d'ajuster la structure des diplômés à la structure de classe de la société ; Boudon (1973) analyse l'école comme le lieu où s'effectuent des choix d'acteurs socialement situés, visant à maintenir ou améliorer leur position sociale d'origine.

Dans ce processus de reproduction sociale, l'évaluation joue un rôle important. Ainsi, déjà en 1966, Bourdieu écrivait : « Pour que soient favorisés les plus favorisés et défavorisés les plus défavorisés, il faut et il suffit que l'école ignore dans le contenu de l'enseignement transmis, dans les méthodes et les techniques de transmission et dans les critères de jugement, les inégalités culturelles entre les enfants des différentes classes sociales : autrement dit, en traitant tous les enseignés, si inégaux soient-ils en fait, comme égaux en droits et en devoirs, le système scolaire est conduit à donner en fait sa sanction aux inégalités initiales devant la culture. » (p. 366).

Dans son livre « *La fabrication de l'excellence scolaire* », Perrenoud (1984) montre comment l'évaluation contribue à fabriquer l'excellence, qui n'est pas une réalité intrinsèque à l'élève, mais est une représentation au même titre que la folie ou la délinquance. Ainsi, l'évaluation scolaire met en évidence certaines différences plutôt que d'autres et, à partir d'inégalités réelles identiques, elle ne fabrique pas les mêmes hiérarchies formelles. Ces mécanismes de sélection et de déformation ne sont pas simplement le fait de l'administration d'épreuves scolaires (le choix des moments de passation, le choix des questions, leur pondération, les critères de correction, l'utilisation des résultats, ...). En effet, d'une part, les évaluations formalisées ne sont

jamais indépendantes des évaluations informelles, implicites, fugitives qui se forment au gré des interactions en classe ; d'autre part, le comportement du maître est influencé par l'évaluation informelle autant que par l'évaluation formelle, en particulier lorsqu'il renvoie à chaque élève une image de sa valeur scolaire. Ainsi très curieusement, l'évaluation tend à rendre plus visible certaines différences qui existaient au départ et en même temps on observe « une certaine indifférence aux différences » (Perrenoud, 1979, p.44) de par l'application d'un enseignement indifférencié à des élèves différents, ce qui contribue à maintenir ou à accentuer les inégalités culturelles.

## **Le paradigme de l'évaluation centrée sur les objectifs**

### ***L'approche tylérienne***

Tyler a développé une approche de l'évaluation essentiellement centrée sur les objectifs. L'évaluation consiste uniquement dans la confrontation entre une performance et des objectifs fixés. Dans cette conception, les objectifs sont déterminés au préalable par les responsables des programmes. L'évaluation a lieu au terme du processus de formation pour mettre en évidence quels élèves maîtrisent quels objectifs. L'information résultante relative aux élèves permet la certification, tandis que celle relative aux objectifs permet aux formateurs de réguler le processus de formation ou, de façon plus globale, aux responsables de réajuster les objectifs et les modalités de mise en œuvre du programme. L'application du modèle tylérien (appelé souvent en France PPO, c'est-à-dire Pédagogie Par Objectifs, ou encore TOP, c'est-à-dire Technique des Objectifs Pédagogiques) suppose huit étapes :

1. la détermination des objectifs,
2. le classement des objectifs dans un système de catégorisation (taxonomie),
3. la définition des objectifs en termes comportementaux,
4. l'établissement des situations et des conditions dans lesquelles la maîtrise des objectifs peut être démontrée,
5. l'explication des buts et des fondements de la stratégie au personnel impliqué dans les situations choisies,
6. le choix et le développement des techniques de mesure appropriées,
7. la collecte des données de performances,
8. la comparaison de celles-ci aux objectifs comportementaux.

Dans l'approche tylérienne, confiance ne sera accordée qu'aux outils pré-construits où l'opérationnalisation aura été soigneusement menée. On recourra donc, selon les objectifs, à des épreuves ou tests d'une part, à des observations critériées en situation provoquée d'autre part.

### ***La pédagogie de la maîtrise***

Développée et promue par Bloom au départ et répandue en Europe grâce aux publications de l'université de Liège (de Landsheere G., 1973, 1980 ; de Landsheere V. et G., 1978), la pédagogie de maîtrise n'est pas le simple prolongement de la conception de Tyler. Il s'agit d'un modèle bien plus riche qui doit beaucoup aux travaux de Carroll, comme l'a très bien montré Scallon (1986). Carroll (1963) a développé un modèle d'apprentissage dont le postulat

de base disait que l'élève réussirait un apprentissage déterminé dans la mesure où il consacrerait le temps qui lui était nécessaire pour apprendre la tâche. Il s'en suit que le niveau d'apprentissage d'un élève est fonction du rapport suivant : le temps « réel » consacré à la tâche par rapport au temps qui lui est nécessaire pour apprendre. Le temps nécessaire dépend lui-même de plusieurs facteurs : l'aptitude spécifique à une tâche donnée ; la capacité de comprendre l'enseignement et la qualité de l'enseignement. Quant au temps réel consacré à la tâche, il dépend essentiellement du temps mis à la disposition de l'élève et de son degré de persévérance, c'est-à-dire du temps qu'il est prêt à consacrer pour réussir la tâche.

Ces apports vont amener Bloom à développer l'idée de pédagogie de maîtrise dont les grandes lignes peuvent se résumer comme suit : préciser clairement les résultats attendus à la fin d'un cours ou d'une séquence d'apprentissage ; préparer les étudiants pour qu'ils puissent entrer avec fruit dans la séquence d'apprentissage ; enrichir l'apprentissage de rétroactions fréquentes et de démarches correctives ; ne pas passer à l'apprentissage ultérieur si l'apprentissage actuel n'est pas suffisamment maîtrisé.

Les fondements de la pédagogie de la maîtrise vont tout naturellement conduire Bloom à distinguer trois types d'évaluation selon les moments et les fonctions de l'évaluation : pour bien préparer l'élève à entrer dans un apprentissage, on recourra à l'évaluation diagnostique ; pour l'aider en cours d'apprentissage, on recourra à l'évaluation formative ; au terme de l'apprentissage, on contrôlera la maîtrise des objectifs fixés par l'évaluation certificative. Cette triple distinction (développée dans un livre célèbre : Bloom et al., 1971) a eu beaucoup de retentissement. De nombreux auteurs l'ont précisée, corrigée (notamment l'ambiguïté du concept d'évaluation diagnostique), complétée (voir par exemple : Cardinet, 1986a ; De Ketele, 1980 ; De Ketele et Roegiers, 1993).

Dans son livre célèbre « *Caractéristiques individuelles et apprentissages scolaires* » (version originale en 1976 ; version française en 1979), Bloom tente de démontrer deux hypothèses : l'historicité de l'individu étant importante pour la réussite de l'apprentissage, il est nécessaire et il suffit d'évaluer dans un premier temps les caractéristiques cognitives et affectives de départ de l'élève ; on peut modifier les caractéristiques de départ et de l'élève (prérequis et motivation), et de l'enseignant ou des deux et ces modifications permettent d'atteindre un apprentissage de plus haut niveau tant chez les individus que dans les groupes. Bloom estime même que l'on peut augmenter ce niveau jusqu'à deux sigmas, ce qui a généré une controverse non encore éteinte.

### ***De nombreuses critiques***

Tout en reconnaissant la contribution de la TOP au développement de l'évaluation, de nombreux auteurs ont mis en évidence un certain nombre de limites, voire de dangers. Ainsi, Hameline (1980) parle de « péché behavioriste » lorsqu'il critique l'exigence de formuler tous les objectifs en termes de comportements observables. De Ketele (1980) stigmatise la tentation du découpage de l'enseignement en micro-objectifs et parle de « pédagogie du saucissonnage ». Charlier (1989) et Donnay et Charlier (1991) mettent en évidence le caractère peu habituel de l'entrée par les objectifs pour les enseignants, qu'il existe d'autres points d'entrée aussi valables et que l'entrée par les objectifs requiert souvent un investissement peu compatible avec le contexte de la classe.

Mais parmi toutes les critiques qui ont été adressées au modèle tylérien et à la pédagogie de la maîtrise, nous retiendrons l'analyse de Barbier (1985). Celui-ci met en évidence la structure pyramidale des rapports de pouvoir de ces modèles : de l'extérieur, les responsables opèrent des choix fondamentaux ; à l'enseignant incombe la confrontation avec l'élève et le jugement final ; l'élève subit le « jugement d'existence » de l'enseignant, particulièrement marqué quand il s'agit du savoir-être. Les critiques ne manquent pas de fuser : accusation de totalitarisme pour les uns, accusation de système reproducteur des inégalités sociales pour les autres, accusation enfin de manquer les valeurs les plus fondamentales qui se laissent moins facilement opérationnaliser comme l'exigent ces modèles.

### **Le paradigme de l'évaluation formative dans un enseignement différencié**

On doit le terme « d'évaluation formative » à Scriven (1967) qui a voulu insister sur le fait que les « erreurs » commises pendant le processus n'étaient ni répréhensibles ni des manifestations pathologiques, mais faisaient partie d'un processus normal d'apprentissage. Ainsi naissaient la distinction et l'opposition entre évaluation sommative (bilan des performances acquises au terme d'un apprentissage) et évaluation formative (évaluation menée pendant un processus d'apprentissage inachevé pour l'améliorer).

La conception de l'évaluation formative va alors se développer dans plusieurs directions. Un des mouvements les plus puissants est celui qui a été généré par le colloque de Genève en 1978 par des chercheurs suisses, belges et français (qui se sont constitués depuis lors en « Association pour le Développement des Méthodologies de l'Évaluation ») qui non seulement ont précisé le concept, mais l'ont orienté vers un paradigme de pédagogie différenciée.

Allal (in Allal, Cardinet et Perrenoud, les éditeurs des actes du colloque de Genève, publiés en 1979) affirme que toute évaluation a une fonction de régulation. Mais si l'évaluation pronostique (celle qui a lieu au début d'un cycle de formation) et l'évaluation sommative (celle qui a lieu à la fin d'une période de formation) régulent dans le sens où elles doivent assurer que les caractéristiques des élèves répondent aux exigences du système, par contre l'évaluation formative doit assurer que les moyens de la formation correspondent aux caractéristiques des élèves.

« Trois étapes essentielles caractérisent donc l'évaluation formative :

- le recueil d'informations concernant les progrès et les difficultés d'apprentissage rencontrées par l'élève ;
- l'interprétation de ces informations dans une perspective à référence critérielle et, dans la mesure du possible, diagnostic des facteurs qui sont à l'origine des difficultés d'apprentissage observées chez l'élève ;
- l'adaptation des activités d'enseignement et d'apprentissage en fonction de l'interprétation faite des informations recueillies. » (p. 132)

Allal a également distingué plusieurs modalités d'application de l'évaluation formative :

- la régulation ou l'évaluation formative rétroactive, c'est-à-dire celle qui consiste à marquer un temps d'arrêt dans les activités d'apprentissage pour établir un constat des objectifs déjà atteints et des objectifs non atteints, afin d'envisager les activités de remédiation pertinentes ;

- la régulation ou l'évaluation formative interactive, c'est-à-dire celle qui s'intègre dans les activités d'enseignement et d'apprentissage : le diagnostic et la guidance individualisées s'opèrent en cours de processus ;

- des modalités mixtes qui peuvent revêtir plusieurs aspects : soit une séquence d'enseignement suivie d'une évaluation formative rétroactive, suivie d'une remédiation s'accompagnant d'une évaluation formative interactive ; soit une séquence d'enseignement s'accompagnant d'une évaluation formative interactive, suivie d'une évaluation formative rétroactive, suivie d'une remédiation ; soit encore un processus continu alternant séquence d'enseignement s'accompagnant d'une évaluation formative interactive et évaluation rétroactive préparant une nouvelle séquence d'enseignement-(ré)-apprentissage.

Plus tard, et par un souci de symétrie évident, émergea également le concept d'évaluation formative proactive, celle qui a pour fonction de recueillir de l'information et de l'interpréter dans le but de préparer une nouvelle séquence d'apprentissage qui tienne compte des caractéristiques des élèves.

La distinction entre évaluation sommative et formative a amorcé un élargissement des conceptions de l'évaluation. Cardinet (1986) en a fait une excellente synthèse autour de neuf questions-clés : quels besoins ?, pour quoi évaluer ?, pour qui évaluer ?, sur quoi évaluer ?, quand évaluer ?, comment recueillir l'information ?, comment interpréter l'information ?, comment utiliser l'information ?, quels sont les problèmes en suspens ?

Au-delà du développement du concept d'évaluation formative, le colloque de Genève a eu le mérite d'amorcer une réflexion très riche autour des concepts d'enseignement différencié et de pédagogie différenciée. Cette réflexion prend son origine dans les travaux issus du paradigme sociologique (cf. supra) et a eu pour but de tenter de répondre aux constats d'inégalités sociales, à « l'indifférence face aux différences » (pour reprendre l'expression de Perrenoud), aux inégalités renforcées par les processus habituels de l'évaluation...

Si l'évaluation formative dans une perspective de différenciation est un élément majeur de la réponse à donner, elle n'en constitue pas moins qu'un élément parmi d'autres : faut-il également faire varier les objectifs pédagogiques en fonction des caractéristiques des élèves ?, faut-il au contraire imposer la maîtrise de compétences minimales identiques et différencier sur les autres compétences (voir l'analyse de de Landsheere, V., 1988) ? faut-il plutôt axer la formation sur des capacités ou compétences transversales de base identiques pour tous mais s'exerçant sur des supports variés et différenciés selon les élèves (Meirieu, 1985, 1987 ; De Ketele, 1982, 1989) ?, faut-il plutôt faire varier les méthodes et mettre l'élève face à plusieurs méthodes afin qu'il trouve celle qui lui convienne (on retrouve ici, notamment, tous ceux qui ont étudié les styles d'apprentissage comme Lamontagne, 1984 et 1987, et de La Garanderie, 1982) ?

Le modèle personnaliste de de Peretti (1981, 1991) mérite d'être signalé, car il ne se centre pas seulement sur la personne de l'élève, mais aussi sur la personne de l'enseignant, lui-même inséré dans un contexte précis avec ses contraintes et ses ressources. Un grand nombre de ses travaux a donc consisté à recueillir non seulement des outils variés (instruments d'évaluation formative, procédures, techniques et méthodes pédagogiques, dispositifs de groupements d'élèves,...), mais aussi, des référentiels et des outils d'analyse

des pratiques pédagogiques. Cette approche « multi-référentielle » et le développement d'une « ingénierie méthodologique » (selon ses propres termes) nous semblent importants dans le cadre d'un paradigme de la différenciation.

## **Le paradigme de l'évaluation au service de la décision**

### ***Le modèle CIPP de Stufflebeam***

Le modèle de Stufflebeam est sans doute le modèle le plus mondialement connu. Contrairement au modèle tylerien, il n'est pas centré sur les objectifs mais sur la décision en situation. Le but de l'évaluation n'est pas de prouver mais d'améliorer, c'est-à-dire prendre des décisions adéquates (« *not to prove, but to improve* », Stufflebeam et Shinkfield, 1985, p. 151). Et ce n'est pas simplement, selon les auteurs, en mesurant l'écart entre performances et objectifs que l'on se met en position de prendre de bonnes décisions.

Le modèle CIPP (Context, Input, Processus, Product) fut donc développé.

L'évaluation du contexte cherche à définir le contexte institutionnel, à identifier les opportunités de répondre aux besoins, à diagnostiquer les problèmes sous-tendant les besoins, à juger si les objectifs proposés permettent de répondre suffisamment aux besoins analysés. Pour mener à bien une telle évaluation, les évaluateurs recourent à des outils variés, tels l'analyse systémique, des enquêtes, des analyses de documents, des entretiens, des tests diagnostiques et la technique Delphi. Une telle évaluation permet de fonder les prises de décision concernant le type de situation à aménager pour la formation, les finalités et les objectifs de la formation, le type d'information à recueillir pour évaluer les résultats.

L'évaluation des inputs (ou « intrants » comme disent les québécois) permet d'identifier les capacités du système, de circonscrire les stratégies alternatives, de prévoir les démarches à implanter et les ressources matérielles, financières et humaines nécessaires. Les évaluations recourent à des outils, telles les réunions d'équipes, la consultation d'inventaires, des visites, des techniques de planification, des études de fiabilité, des simulations. L'évaluation permettra de fonder les prises de décision concernant le choix des ressources et des stratégies ainsi que le type d'informations à recueillir pour évaluer la façon dont la stratégie prévue a été implantée.

L'évaluation du processus vise à identifier et même prédire dans certains cas les dysfonctionnements par rapport au dispositif prévu, à fournir les informations nécessaires pour les décisions préprogrammées et à prendre pendant le processus de formation, enfin à enregistrer et à porter un jugement sur les événements et activités de formation. Carnets de bord, méthodes des incidents critiques, réunion d'équipes, entretiens, analyse des représentations des acteurs de la formation, analyse des productions seront les outils privilégiés d'une telle évaluation. Celle-ci permet de fonder les prises de décision concernant les étapes prévues, les modifications à entreprendre et le type d'informations à recueillir et à conserver pour l'interprétation des résultats de la formation.

L'évaluation du produit consiste à rassembler les descriptions et jugements concernant les résultats, à les mettre en relation avec les objectifs, le

contexte, les inputs et le processus, enfin à les interpréter en termes de jugement de valeur. L'évaluateur recourra à deux types d'outils, les outils préconisés par le modèle tylerien mais aussi toutes formes d'outils susceptibles de récolter les jugements des acteurs concernant les résultats de la formation. Ces deux types d'informations seront soumis à une analyse — non seulement quantitative — mais aussi qualitative. Le rapport final d'évaluation consignera les jugements de valeur en distinguant les effets positifs et négatifs par rapport aux résultats prévus, voire les résultats imprévus.

Un des grands mérites du modèle de Stufflebeam est d'être un modèle global. Mais c'est aussi sa principale limite. Le modèle est essentiellement pensé en termes d'évaluation globale des formations et non en termes d'évaluation des agents en formation, comme le font les enseignants en classe (cf. la distinction de Barbier, 1985). Cependant, bien des composantes du modèle peuvent inspirer judicieusement l'évaluation qui se déroule dans le monde microscopique de la classe.

### ***Le modèle de l'évaluation au service d'une pédagogie de l'intégration selon De Ketele***

Ce modèle répond à une double préoccupation : une application du modèle de Stufflebeam (fort large, puisque conçu au départ pour l'évaluation des programmes de formation) au cadre plus restreint de l'évaluation scolaire ; un développement de ce modèle en y intégrant le paradigme d'une pédagogie de l'intégration.

Appliqué à l'évaluation scolaire, le modèle de Stufflebeam implique que toute évaluation menée autour des apprentissages a pour fonction principale de prendre une décision d'action (« *Evaluer pour mieux agir* », selon l'expression d'Hadji, 1992), et non simplement de faire un constat, à plus forte raison un jugement. Si le processus évaluatif comprend des constats et des jugements à certaines étapes, ceux-ci n'ont de valeur qu'en tant qu'éléments concourant à la prise de décision. Nous pouvons donc résumer le processus évaluatif dans la définition suivante : « Evaluer consiste à recueillir un ensemble d'informations reconnues comme suffisamment pertinentes, valides et fiables, et à examiner le degré d'adéquation entre cet ensemble d'informations et un ensemble de critères jugés suffisamment adéquats aux objectifs fixés au départ ou ajustés en cours de route, en vue de fonder une prise de décision. » (De Ketele, 1980 ; De Ketele et Roegiers, 1993).

Cette définition met en évidence un certain nombre d'opérations dans le processus évaluatif, dont certaines précèdent nécessairement d'autres. La première n'est certainement pas (contrairement à beaucoup de pratiques) de recueillir des informations, mais réside dans le fait qu'il faille d'abord se poser la question du « pour quoi » évaluer, c'est-à-dire la fonction de l'évaluation (« Quel type de décision serai-je amené à prendre au terme de l'évaluation ? »). Toutes les opérations suivantes seront conditionnées par la réponse à cette première question (qui peut en corrélation en entraîner d'autres, comme par exemple : « pour qui ? », ...). En fonction du type de décision à fonder, l'évaluateur est amené à se poser la question des objectifs qu'il convient d'évaluer. Contrairement au modèle de Tyler, ces objectifs ne sont pas toujours et nécessairement prédéterminés, ils peuvent être révélés en cours de route ou résulter d'ajustements décidés en cours d'apprentissage. L'évaluation des objectifs suppose la définition de critères adéquats, voire

d'indicateurs lorsque cela est possible et pertinent. C'est seulement sur cette base que peuvent être choisies les situations qui permettront de recueillir l'information jugée nécessaire. Commencent alors les opérations concrètes de recueil, de traitement, d'interprétation et de synthèse de l'information qui conduiront à la prise de décision. Si ces opérations ne sont pas toujours aussi linéaires que ce bref exposé ne le laisse sous-entendre (en effet, des ajustements sont souvent jugés nécessaires qui impliquent des retours en arrière), une hiérarchie existe entre ces opérations, dans le sens que la qualité d'une opération dépend de celles qui la précèdent.

De Ketele et Roegiers (1993, chap. 2) ont développé et précisé les qualités attendues d'une évaluation autour de trois concepts fondamentaux : la pertinence, la validité et la fiabilité. La pertinence signifie que l'évaluation répond à sa fonction première. La validité suppose qu'on évalue réellement ce que l'on déclare évaluer. La fiabilité a trait à la confiance que l'on peut avoir dans les opérations effectuées.

Ces qualités peuvent être examinées à un macro- ou à un micro-niveau, autour des questions suivantes :

<b>Macro (évaluation)</b>	
Macro-pertinence	"Est-ce que je ne me trompe pas d'objectif ?"
Macro-validité	"Mes critères permettent-ils de vérifier ce que je déclare vouloir vérifier ?"
<b>Micro (recueil d'informations)</b>	
Micro-pertinence	"Est-ce que je ne me trompe pas d'informations à recueillir ?"
Micro-validité	"La stratégie mise en place me donne-t-elle toutes les garanties que l'information que je vais recueillir est bien celle que je déclare vouloir recueillir ?"
Micro-fiabilité	"La façon de recueillir l'information est-elle semblable d'une personne à l'autre, d'un endroit à l'autre, d'un moment à l'autre ?"
Macro-fiabilité	"L'utilisation que je fais des critères est-elle la même pour tout le monde ?"
<b>(communication)</b>	
Pertinence	"Est-ce que je ne me trompe pas de résultats à communiquer ?"
Validité	"Est-ce que les résultats que je communique sont bien ceux que je déclare vouloir communiquer ?"
Fiabilité	"Est-ce que le décideur peut se fier à ces résultats ?"

Un des intérêts du modèle de Stufflebeam réside dans la quadruple distinction « contexte, entrées, processus, sorties » et dans l'articulation entre ces quatre composantes. A un niveau plus microscopique, la classe et même l'élève sont des systèmes ; en conséquence, cette quadruple distinction s'applique et on peut penser que pour chacune d'elle correspondent une ou plusieurs fonctions différentes de l'évaluation, telles l'évaluation des besoins, l'évaluation pronostique, l'évaluation formative, l'évaluation certificative.

Mais l'apport le plus important au modèle de Stufflebeam, appliqué au contexte de la classe, réside dans le développement d'une pédagogie de l'intégration.

La pédagogie de l'intégration est une tentative pour lutter contre une pédagogie du « saucissonnage » habituellement pratiquée dans le cadre scolaire : l'école n'est souvent qu'une juxtaposition de cubes les uns à côté des autres ou les uns au-dessus des autres ; les horaires sont des tranches de temps homogènes que l'on utilise le plus souvent indépendamment les uns des autres ; les programmes juxtaposent des disciplines les unes à côté des autres ; un programme est trop souvent une suite de chapitres indépendants ; un chapitre est lui-même souvent une suite de leçons...

Même l'évaluation n'échappe pas au piège du saucissonnage. Le terme d'évaluation sommative est révélateur. Les spécialistes de l'évaluation eux-mêmes encouragent cette tendance lorsqu'ils disent qu'une bonne évaluation est un échantillon de questions représentatives de l'univers des objectifs visés ou/et des contenus de l'apprentissage.

La pédagogie de l'intégration dit au contraire que le tout n'est pas la somme des parties. Ce n'est pas parce qu'un élève réussit toutes les questions d'un test qu'il a « compris » (prendre ensemble) et qu'il sait utiliser ce qu'il a appris dans un contexte où il ne suffit pas de juxtaposer les acquis, mais de faire des choix et de nouveaux assemblages avec les objets des apprentissages antérieurs.

C'est pourquoi nous avons introduit les nouveaux concepts d'objectifs terminal ou intermédiaire d'intégration (De Ketele, 1980, De Ketele et alii, 1989). Un objectif d'intégration possède les caractéristiques suivantes : (a) la compétence visée s'exerce sur une situation d'intégration, c'est-à-dire une situation complexe comprenant de l'information essentielle et de l'information parasite et mettant en jeu les apprentissages antérieurs les plus significatifs ; (b) la compétence est une activité complexe nécessitant l'intégration et non la juxtaposition des savoirs, savoir-faire et savoir-être appris antérieurement ; (c) la situation d'intégration est la plus proche possible des situations naturelles auxquelles seront confrontés les élèves plus tard ; (d) la compétence est orientée vers le développement de l'autonomie et donc le « savoir-devenir ».

Préparer un cours au début de l'année réside essentiellement à se fixer un bon objectif terminal d'intégration qui comprend les apprentissages les plus importants et, à partir de là, quelques objectifs intermédiaires d'intégration qui seront des passages obligés dans la progression et dans l'apprentissage de cette compétence fondamentale qu'est le « savoir intégrer ».

Au niveau de l'évaluation, il en résulte que l'évaluation certificative finale prend comme objet l'objectif terminal d'intégration, en prenant bien soin de se définir des critères et des indicateurs minimaux par rapport à des critères et indicateurs que nous pourrions qualifier de perfectionnement.

En cours d'année, des évaluations-bilans à visée formative porteront sur les objectifs intermédiaires d'intégration et, si le besoin s'en fait sentir, des évaluations diagnostiques en profondeur porteront sur des aspects plus spécifiques de l'apprentissage (voir De Ketele et Paquay, in Allal, 1991).

### **Le paradigme de l'évaluation centrée sur le consommateur selon Scriven (1967, 1983)**

De Scriven, les pédagogues n'ont malheureusement retenu trop souvent que la seule distinction entre évaluation sommative et formative. L'approche de Scriven est cependant originale et innovante sur bien des points : (a) il

préfère l'évaluation en amateur (« amateur evaluation ») à l'évaluation professionnelle (« professional evaluation ») ; (b) il insiste sur l'importance d'évaluer les effets de la formation et de l'évaluation sur les « clients » ; (c) il milite pour une évaluation dégagée des objectifs (« goal free evaluation »), c'est-à-dire centrée sur tous les effets de la formation ; (d) il défend l'idée que le concept-clé de l'évaluation est le concept de jugement de valeur (et non les concepts d'objectif ou de décision) ; (e) comme seule base de jugement de valeur, il ne reconnaît que l'analyse des valeurs et des besoins des consommateurs (ce qui s'écarte très nettement de la position de Tyler et même de celle de Stufflebeam qui, s'il considère comme important dans l'évaluation du contexte la prise en compte des besoins du public-cible, reconnaît une primauté aux décideurs ; par contre, on se rapproche beaucoup plus de la pensée de Barbier, 1985) ; (f) il souligne la nécessité de comprendre (d'« illuminer ») les effets à la lumière de ce qui s'est passé dans la formation ; (g) pour lui, le jugement final consiste à comparer la valeur ou le mérite relatif des objets soumis à l'évaluation (stratégies de formation mises en œuvre, stratégies alternatives, personnes en formation, ..., selon les cas) ; (h) enfin, il prône la nécessité d'une méta-évaluation, c'est-à-dire d'une évaluation de l'évaluation par des personnes externes compétentes, si possible avant l'implantation de l'évaluation et avant la diffusion du rapport final.

Ainsi donc, le modèle de Scriven signifie que l'évaluateur doit partir l'esprit libre de tout objectif. Sa première tâche consiste à rassembler un maximum d'informations par des outils variés sur toutes les composantes du système de formation. Sa dernière tâche est de prononcer un jugement de valeur.

Mais comment passer du recueil d'informations au jugement de valeur ? Par l'analyse des besoins réels des consommateurs qui fourniront les critères du jugement de valeur (les travaux de Figari sur les référentiels vont dans ce sens).

Mais qui sont les consommateurs ? Scriven distingue deux types de consommateurs : la clientèle ou marché-cible qui recevront les personnes formées et les « utiliseront » d'une part ; les vrais consommateurs d'autre part, c'est-à-dire les récipiendaires de la formation. Dans l'analyse des besoins du marché-cible, l'évaluateur se posera des questions, telles par exemple : quelles sont les compétences et attitudes que les élèves sortant du lycée doivent posséder et sur lesquelles la communauté universitaire doit pouvoir tableer pour assurer une formation universitaire de qualité ? Quels sont les savoir-faire et les savoir-être dont les élèves-maîtres doivent faire preuve au terme de leur formation pour répondre aux exigences de la fonction enseignante telle qu'elle est vue par les chefs d'établissement, les enseignants, les parents, les élèves ?... Dans l'analyse des besoins du « consommateur vrai » qu'est l'élève de terminale, l'évaluateur tentera d'identifier les besoins réels de celui-ci (comme par exemple, un concept de soi positif, une bonne préparation au bac, ...). L'analyse des besoins du futur enseignant révélera peut-être un besoin d'être sécurisé quant à la façon de conduire sa classe... Les deux types de besoins (ceux du marché-cible et ceux du récipiendaire de la formation) ne risquent-ils pas parfois de rentrer en conflit ? Le seul élément de réponse que nous découvrons chez Scriven réside dans le fait que l'analyse des besoins doit prendre en considération les besoins « réels » et non les besoins perçus comme idéaux.

Centré sur les besoins des consommateurs, le modèle de Scriven reconnaît donc une place à l'évaluation sommative. Elle consistera par exemple à

comparer plusieurs stratégies de formation par rapport aux besoins réels de tels consommateurs ; ou elle comparera les savoir-faire et les savoir-être de plusieurs futurs enseignants par rapport aux besoins réels des écoles, voire d'une école particulière. On comprend aussi que, pour des raisons d'objectivité, Scriven défend l'idée d'un évaluateur externe pour l'évaluation sommative ou certificative.

### **Le paradigme de l'évaluation centrée sur le client ou paradigme de l'évaluation répondante selon Stake (1967, 1975, 1976)**

L'itinéraire biographique de Stake est un itinéraire curieux et intéressant. Initialement professeur de mathématique, il présentera une thèse doctorale en psychométrie. Il devint de plus en plus sceptique vis-à-vis de la théorie classique de la mesure pour résoudre les problèmes de l'évaluation scolaire et vis-à-vis des tests en général pour comparer les stratégies différentes de formation. Il tentera alors de définir un modèle d'évaluation des curricula qui tienne compte des apports jugés positifs de Tyler, Stufflebeam et Scriven, mais qui en dépasse les limites.

S'il ne nie pas l'intérêt de comparer les résultats observés aux résultats attendus comme le voulait Tyler, il estime cela insuffisant. En effet, il déclare que les résultats attendus et même les résultats observés ne sont pas nécessairement les mêmes selon les personnes, c'est-à-dire selon les différents types de clients ou personnes impliquées dans des situations différentes de prise de décision. Il développe alors une approche centrée sur le client ou plus exactement sur les clients, car, dans toute évaluation, dira-t-il, les clients ne sont pas seulement ceux qui demandent l'évaluation, mais aussi toutes les personnes impliquées (c'est-à-dire ce qu'il appelle les « audiences »).

S'il admet avec Scriven que toute évaluation comporte un jugement de valeur, il trouve vain de porter un jugement final comparatif (et de ce fait refuse l'évaluation sommative). Pour cela, dit-il, il faudrait que le jugement puisse s'appuyer sur des normes de jugement uniques. Or, les normes de jugement sont multiples. Le rôle de l'évaluateur est donc pour lui : (a) de récolter le plus objectivement possible les informations, les différentes normes de jugement et les différents jugements des différentes personnes ; (b) d'en analyser la congruence ; (c) de fournir par là une meilleure compréhension de la situation en vue de permettre aux différentes audiences de mieux comprendre et d'améliorer le processus de formation (d'où le concept d'évaluation répondante).

Stake tente alors de concilier d'une part la pensée de Stufflebeam qui insiste sur la nécessité d'évaluer le contexte, les inputs, le processus et les produits et d'autre part la pensée de Scriven, centrée sur le concept de jugement de valeur et de critères relatifs aux besoins des consommateurs. Il propose ce qu'il appelle un « format pour la collecte des données ».

Ce format comprend trois composantes. La première consiste à décrire les « fondements du programme » de formation. La seconde consiste à remplir une « matrice de description », comprenant deux colonnes (d'une part, les « intentions », c'est-à-dire ce que les différentes personnes veulent ; d'autre part, les « observations », c'est-à-dire ce que les différentes personnes perçoivent) et trois lignes (les « antécédents », c'est-à-dire les intentions et les observations avant la mise en route du programme ; les « transactions », c'est-à-dire les intentions et les observations qui se sont manifestées en

cours de processus de formation ; les « résultats », c'est-à-dire les intentions et les observations résultant de la formation). La troisième consiste à établir la « matrice de jugement », comprenant également deux colonnes (les « critères », c'est-à-dire ici ce que les personnes en formation doivent apprendre selon les différentes personnes ; les « jugements », c'est-à-dire la valeur attribuée par les différentes personnes aux composantes du programme) et également trois lignes (à nouveau les « antécédents », les « transactions » et les « résultats », mais en termes de « critères » au sens de performances à apprendre et de « jugements » au sens de valeur attribuée).

Le travail de l'évaluateur consistera donc d'abord à produire le plus honnêtement possible les informations objectives et subjectives des différentes facettes de ce format et relatives aux divers clients impliqués. Ensuite, il mènera une double analyse : une analyse en terme de congruence et une analyse en terme de contingence logique.

L'analyse en terme de congruence désigne l'analyse des accords et désaccords : (a) entre les différentes audiences ; (b) entre intentions et observations ; (c) entre critères et jugements ; (d) entre descriptions et jugements ; (e) entre fondements du programme, descriptions et jugements.

L'analyse en terme de contingence logique désigne le caractère logique du passage : (a) entre antécédents et transactions ; (b) entre transactions et résultats ; (c) entre les fondements du programme et la succession antécédents, transactions et résultats.

Enfin, les résultats de cette double analyse sont synthétisés et communiqués aux différents clients sous des formes et des langages appropriés, de telle sorte que chaque public concerné puisse apporter sa part à l'amélioration des programmes.

Avec Stufflebeam et Shinkfiels (1985, p. 66), on peut cependant reprocher au modèle de Stake de proposer des outils d'évaluation insuffisamment développés ou validés (comme, par exemple, les études de cas, les rapports contradictoires, les sociodrammes, ...), de produire une information dont la fiabilité et l'objectivité sont souvent sujettes à caution, de travailler avec des évaluateurs peu formés et souvent trop impliqués (la méta-évaluation proposée par Scriven pourrait cependant être considérée comme une parade), de fournir des conclusions souvent équivoques et contradictoires, donc souvent peu utiles, et même de travailler dans un contexte où la manipulation est facilement possible. A ces objections, nous en ajouterons une plus terre à terre : comment la rendre possible et réaliste dans le contexte de la pratique scolaire habituelle ?

### **Le paradigme économique**

On peut situer le grand développement du paradigme économique à partir du concept de « rendement de compte » (« accountability ») lancé par le président Nixon et son administration. Des sommes considérables vont être allouées aux recherches visant à évaluer l'efficacité des systèmes d'enseignement (pensons aux recherches de l'IEA). Selon Bressoux (1993), ce mouvement s'est dessiné en réaction aux conclusions des travaux de Coleman et alii (1966) et Jencks et alii (1972) qui déclaraient que l'école n'exerce pas (ou très peu) d'effets spécifiques sur les acquisitions des élèves : celles-ci sont bien davantage déterminées par les caractéristiques individuelles des élèves

et notamment par leur origine socio-économique. Indépendamment, aux États-Unis, deux courants de recherche se sont développés dans les années 70 : l'un, dans une perspective militante, s'inscrit dans une lutte contre les inégalités sociales et étudie les effets-écoles ; l'autre étudie les répercussions de l'enseignement des maîtres sur les comportements et les acquisitions des élèves et prend donc la classe comme unité d'observation. Ces deux courants ont été relayés en France par l'Institut de Recherche en Economie de l'Éducation à Dijon (IREDU).

Au centre des préoccupations du paradigme économique se trouvent le constat de la variabilité des « acquisitions des élèves » et le désir d'identifier les déterminants de cette variance, principalement les effets liés au maître, à la classe et à l'école, une fois éliminés les effets liés aux caractéristiques individuelles, surtout celles relevant du niveau socio-économique.

Les travaux de l'IREDU se centrent principalement autour de deux variables dépendantes : l'efficacité et l'équité.

Le concept d'efficacité est lié au concept de progression des acquisitions des élèves : un enseignant ou un programme seraient d'autant plus efficaces avec un public donné qu'il serait davantage capable de faire progresser les élèves dans un laps de temps déterminé. Méthodologiquement parlant, cela suppose au moins deux mesures comparables dans le temps des performances acquises par les élèves. Opérationnellement parlant, l'efficacité n'est cependant pas mesurée par la simple différence arithmétique entre résultat final et résultat initial (qui ne tient pas compte des biais liés aux effets de « plafond » ou de « plancher » d'une part, et au fait que la différence arithmétique est généralement corrélée avec le pré-test d'autre part), mais est estimée à partir des résidus d'une régression du score final sur le score initial : les résidus positifs (ceux qui se situent au-dessus de la droite de régression) exprimeront une efficacité plus grande alors que les résidus négatifs (ceux qui se situent en-dessous de la droite de régression) traduiront une régression relative et donc une efficacité moins grande.

Plus philosophique ou moral, le concept d'équité est défini d'un point de vue économique comme le fait que le fossé constaté au départ entre les faibles et les forts n'augmente pas, voire même diminue. D'un point de vue statistique, cela signifie que la situation d'équité maximale théorique correspond au cas où les résultats aux épreuves finales ne dépendent pas de ceux obtenus aux épreuves initiales. Opérationnellement, les chercheurs de l'IREDU la mesurent par la pente de la régression linéaire simple du résultat final sur le résultat initial. On peut trouver une explication simple du modèle de l'IREDU dans la thèse de Bréssoux (1993).

Ces deux variables dépendantes étant définies, la préoccupation des chercheurs fonctionnant avec ce paradigme est d'essayer de déterminer, « toutes choses restant égales par ailleurs », quels sont les facteurs qui influencent le plus l'efficacité et l'équité. Dans ce but, l'IREDU propose un modèle méthodologique original que nous qualifierons de « hiérarchique progressif », qui consiste à isoler d'abord la variance expliquée par les caractéristiques individuelles de l'élève (et surtout le poids des caractéristiques socio-économiques), pour ensuite isoler dans la variance non expliquée la part de variance expliquable par l'effet-maître, puis la part de l'effet-classe indépendante des effets-maîtres, enfin la part de l'effet-école indépendant des effets précédents. La part de variance une fois estimée et expliquée par un effet, les chercheurs de l'IREDU tentent de voir quelles sont les variables

qui ont tendance à composer l'effet en question, par des approches quantitatives ou qualitatives (Duru-Bellat, 1991).

Le paradigme économique est évidemment un paradigme séduisant pour les macro-décideurs qui ont la tâche et le désir de gérer au mieux les deniers de la collectivité. Les résultats des travaux de l'IREDU (Mingat, 1983, 1984, 1987, 1991 ; Duru-Bellat et Mingat, 1985, 1988 ; Duru-Bellat et Leroy-Audouin, 1990) semblent indiquer que : (a) l'effet-maître est aussi important que les effets liés aux caractéristiques individuelles de l'élève ; (b) du moins à l'école primaire, l'effet-école serait pour l'essentiel une aggrégation des effets-maîtres : l'effet spécifique de l'école serait donc faible ; (c) les attentes des maîtres vis-à-vis de leurs élèves joueraient un rôle important dans l'effet-maître ; (d) les maîtres les plus équitables sont souvent aussi les plus performants, mais d'autres cas de figure existent.

Le paradigme économique est essentiellement une macro-évaluation : c'est sa principale qualité et sa principale limite. En tant que macro-évaluation, il exige le déploiement de gros moyens (échantillons importants et minutieusement constitués, arsenal méthodologique complexe, ...). Il demande en outre des instruments de mesure valides et fiables. On ne s'étonnera donc pas que les variables dépendantes mesurées soient des acquisitions scolaires classiques (la lecture, le calcul,...). Il est d'autres variables importantes que celles-là, comme la curiosité, la créativité, l'indépendance, l'attitude vis-à-vis de l'école, l'adaptation ..., plus difficilement évaluables et qui cependant peuvent faire des différences importantes entre maîtres (cf. les 153 recherches analysées par Giaconia et Hedges, 1982) et probablement seraient davantage liés à l'effet-école.

Si le paradigme économique est, et doit rester, un modèle macro au service de la gestion éloignée (à condition qu'il soit complété par d'autres approches), il n'en contient pas moins deux concepts-clés importants également pour la gestion quotidienne de la classe : l'efficacité conçue comme le fait de faire progresser les acquisitions des élèves ; l'équité conçue comme le fait de ne pas accroître et même de diminuer l'écart qui existe au départ entre les forts et les faibles.

### **Le paradigme de l'évaluation comme processus de régulation**

Selon Allal (in Allal, Cardinet, Perrenoud, 1979), la régulation est le concept central de l'évaluation, dans le sens où toute évaluation — explicitement ou implicitement — est un moyen de régulation à l'intérieur d'un système de formation. Cela reste vrai, que l'on se place au niveau macro- ou micro-scopique.

Le paradigme de l'évaluation comme processus de régulation, développé actuellement par de nombreux chercheurs, pourrait être un paradigme fédérateur, tout en permettant à des « écoles » différentes de conserver leur identité. C'est ainsi que nous pouvons interpréter la classification proposée par Bonniol, Genthon et Roger (1989). Selon eux, on peut distinguer les évaluateurs qui privilégient : (a) « le bilan et la correction des produits » (« référentiel de contrôle », selon l'expression de Figari, 1993) ; (b) « la facilitation des apprentissages » (« référentiel de compétences cognitives ») ; (c) « l'aide à la construction de l'identité individuelle et collective » (« référentiel de représentations sociales ») ; (d) « la régulation du fonctionnement individuel et collectif » (« référentiel d'interactions »), etc.

A la lecture de cette classification, on devine les fondements systémiques du paradigme de la régulation. Comme le montre Monteil (1985), la formation est un système, lui-même composé de sous-systèmes en interactions : à l'entrée, un système composé de l'environnement psychosocial ; le groupe en formation ou le système régulé ; le système supportant la formation ou système régulateur ; le produit ou le système d'application.. Toujours selon Monteil, « le paradigme systémique remet en cause deux préceptes cartésiens : la causalité (à laquelle le systémique oppose la finalité) et le réductionnisme (auquel le systémique oppose la globalité) ».

Il y a cependant de multiples façons de faire fonctionner le paradigme systémique et, à travers lui, le paradigme de la régulation. Ou bien on le fait fonctionner comme un modèle cybernétique et on se trouve dans un système fermé qui privilégie le « contrôle de qualité », sur la base d'un référentiel prédéterminé. Ou bien on le fait fonctionner de façon ouverte, « à la recherche de sens » (Ardoine et Berger, 1989), voire à la création d'un sens nouveau ». De nombreuses modalités de fonctionnements plus ou moins ouverts peuvent être envisagées, car les sous-systèmes interagissent avec des poids différents selon les lieux et les moments. En effet, comme le montre Figari (1993), les référentiels sont multiples à l'intérieur du système de formation et rentrent souvent en compétition quand ils ne s'ignorent pas : le référentiel de l'administrateur qui est un référentiel de gestion, celui de l'inspecteur qui est politique, celui du consultant ou de l'auditeur qui est d'expertise, celui du chercheur qui est scientifique, celui de l'enseignant qui est pédagogique, celui de l'élève qui lui est centré sur les représentations de la réussite et de l'échec, celui des parents centré sur un devenir, celui du chef d'établissement centré sur la gestion des ressources humaines et matérielles... Ainsi selon les cas, les approches, les lieux et les moments, la visée privilégiera tantôt la recherche de sens, tantôt l'action ; et la régulation sera tantôt davantage du contrôle, tantôt davantage une stratégie innovante.

Nous sommes donc bien dans la complexité et la mouvance. Le paradigme est prometteur car riche dans ses fondements, mais il reste encore à construire. Certes, des synthèses existent déjà. Nous pensons particulièrement aux travaux de d'Hainaut, et principalement la synthèse qu'il a publiée pour l'UNESCO « La régulation dans les systèmes éducatifs. Guide méthodologique » (1980b). Cette synthèse est une des plus complètes et des plus opérationnelles, mais elle prescrit comme point d'entrée la politique éducative dont la responsabilité incombe aux politiques ; or des points d'entrée multiples existent et qu'il convient de prendre en compte et d'articuler. Il reste donc à construire un modèle de la régulation qui, tant théoriquement qu'opérationnellement, articule la multiréférentialité et puisse prendre en compte différents degrés d'ouverture ou de fermeture du système dans ses différents niveaux macro-, méso- et micro-scopiques.

#### **En guise de conclusion provisoire : l'évaluation à la question, l'évaluation en question(s)**

Avec son équipe du CEPEC, Delorme publiait en 1987 un livre au titre évocateur « L'évaluation en questions ». Dans cette conclusion, nous pourrions aussi mettre en évidence un certain nombre de questions que pose l'état actuel des travaux sur l'évaluation ; mais nous pourrions aussi les soumettre à la question et au pouvoir des juges et, peut-être, les remettre en question !

Dans certains pays, l'évaluation est de plus en plus souvent assignée en justice. Dans son livre « Faire réussir. Faire échouer », Viviane De Landsheere (1988) raconte le procès retentissant qui a opposé en 1977 dix étudiants noirs au Département de l'Éducation de Floride. Ils s'étaient vu refuser un certificat d'alphabétisation fonctionnelle au terme de l'enseignement secondaire après avoir été soumis à trois reprises à un test de compétences minimales. 20 % d'élèves noirs étaient dans ce cas, pour seulement 1,9 % d'élèves blancs. Sur cette base, ils intentèrent avec succès une action en justice pour violation du droit constitutionnel de tous les citoyens à jouir d'une « égalité de traitement ». Un passage du jugement rendu par les juges est particulièrement éloquent :

« 1. Les élèves doivent être informés des objectifs spécifiques sur lesquels le test (ou l'examen) décidant de l'attribution du diplôme portera. Cette information doit être donnée au moment où l'enseignement porte sur ces objectifs.

2. Le curriculum proposé aux élèves doit comprendre l'enseignement correspondant aux objectifs couverts par l'examen.

3. L'acquisition des habiletés est un processus cumulatif. L'enseignement donné aux élèves doit se faire en séquences rationnelles et ordonnées permettant d'acquérir la compétence au cours d'un processus de développement approprié.

4. La quantité de temps consacré à l'enseignement d'une habileté ou d'une unité de connaissances particulières est importante.

5. Le moment où l'enseignement se fait aux élèves est important. Non seulement ils doivent se voir enseigner une habileté pendant un temps suffisant, mais ils doivent aussi recevoir un enseignement ou faire une révision des enseignements antérieurs juste avant l'administration du test.

6. Le processus d'enseignement doit comprendre un mécanisme permettant de vérifier si les objectifs poursuivis sont effectivement atteints par les élèves, car l'enseignement et l'apprentissage ne se succèdent pas toujours immédiatement.

7. Les élèves doivent se voir proposer un enseignement de rattrapage s'ils n'ont pas maîtrisé minimalement une compétence. »

(De Landsheere, 1988, p. 222)

Ce jugement est éloquent à plus d'un titre. Il met en évidence la responsabilité civile du formateur et de l'évaluateur (« accountability »). Mais il importe de noter ici que le jugement se base sur le paradigme rationnel de la pédagogie de maîtrise (cf. supra), certes dominant à l'époque dans l'intelligentsia des experts américains en pédagogie. Il serait intéressant d'imaginer d'autres procès qui se fonderaient sur d'autres paradigmes, comme le paradigme docimologique (exemple : la contestation de la note donnée au bac par des correcteurs différents !) ou comme le paradigme de la régulation (exemple : l'accusation d'une régulation insuffisante à l'intérieur d'un établissement scolaire !)... On voit ici le poids des référentiels à la base des paradigmes.

Dans les paradigmes que nous avons passés en revue, on assiste à des conceptions de l'évaluation qui offrent des degrés d'élargissement différents. Ainsi, le paradigme intuitif ou celui de la pédagogie de maîtrise ou celui de l'évaluation formative sont nettement plus étroits que ceux de Stufflebeam, de Scriven ou de Stake. On peut se demander si un système de formation n'est pas caractérisé par un ensemble de paradigmes (et donc de référentiels) plus

ou moins indépendants et si un des problèmes majeurs pour les travaux à venir n'est pas de les articuler théoriquement et opérationnellement. Comme nous l'avons souligné déjà, le paradigme de la régulation nous semble une voie intéressante pour répondre à cette question. Certaines pistes ont été ébauchées dans les ateliers, notamment celui sur « L'évaluation des apprentissages scolaires », que le réseau francophone « Recherche en Education et Formation » a tenu à Sherbrooke en octobre 1992 (les actes paraîtront en 1993).

Mais au terme de cette synthèse (forcément partielle et partielle), nous voudrions penser aux praticiens de l'évaluation et à tous ceux qui sont confrontés aux exigences de la pratique (les nombreux formateurs-chercheurs d'institutions, telles les MAFPEN, l'INRP, l'INRAP, le CEPEC, ..., les associations d'enseignants...). Il ne faudrait pas oublier leurs préoccupations et leurs apports. Pour leur rendre hommage, je voudrais rapporter ce qu'une enseignante disait au terme d'une formation à l'évaluation : « Appliquée à la lettre, l'évaluation telle qu'il (l'animateur de la formation) la préconise requiert une énergie, une disponibilité, une puissance de travail, une imagination en même temps qu'une rigueur logique assez peu communes ! Je m'essouffle rien qu'à lire ces écrits qui ne provoquent pas toujours sur moi l'effet stimulant souhaité par l'auteur. Il m'arrive de me sentir accablée. Faut-il vraiment faire tout cela pour évaluer valablement ? Devant l'énormité de la tâche, la tentation me vient de laisser tomber les bras ou de prêter une oreille bienveillante aux chants de sirène de Rogers. » Mais l'enseignante se reprend et ajoute : « Le mauvais esprit que je suis ne peut se défendre d'une idée gênante : Et si l'attitude de Rogers était une dérobade ? Une démission ? Les rapports humains sont tellement plus agréables quand on peut dispenser de juger !... Ne le dit-il pas lui-même ? (...) A la différence de Rogers, je crois beaucoup à la paresse naturelle de l'homme ! S'il n'est pas confronté à des épreuves, je doute qu'il devienne naturellement **chercheur actif**. » (Moussiaux, mémoire inédit).

Les experts en évaluation, le plus souvent, « enseignent » l'évaluation à leurs étudiants. Puissent-ils interroger leur pratique évaluative et éviter la tentation schizophrénique qui consiste à faire des écrits et des pratiques deux mondes différents !

Jean-Marie De Ketele  
Université de Louvain

Laboratoire de Pédagogie expérimentale, Louvain-la-Neuve (Belgique)

### Bibliographie

ALLAL P. (1991). — **Vers une pratique de l'évaluation formative. Matériel de formation continue des enseignants**. Bruxelles, De Boeck.

ALLAL A., CARDINET J. et PERRENOUD P. (Ed.) (1979). — **L'évaluation formative dans un enseignement différencié**. Berne, Peter Lang.

ARDOINO J. et BERGER G. (1986). — L'évaluation, comme interprétation, **Pour**, 107, Toulouse, Privat.

ARDOINO J. et BERGER G. (1989). — **D'une évaluation en miettes à une évaluation en actes**. Paris, Andsha-Matrice.

ARDOINO J. et BERGER G. (1989b). — Fondements de l'évaluation et démarche critique, **Bulletin de l'AECSE**, 6, pp. 3-11.

BARBIER J.M. (1985). — **L'évaluation en formation**. Paris, PUF.

- BAUDELLOT C. et ESTABLET R. (1971). — **L'école capitaliste en France**. Paris, Maspéro.
- BAUDELLOT C. et ESTABLET R. (1975). — **L'école primaire divisée**. Paris, Maspéro.
- BLOOM B.S., et al. (1971). — **Handbook on formative and summative evaluation of student learning**. New-York, Mac Graw Hill.
- BLOOM B.S. (1979). — **Caractéristiques individuelles et apprentissages scolaires**. Bruxelles, Labor ; Paris, Nathan.
- BLOOM B.S. (1976). — **Human characteristics and School Learning**. New-York, Mac Graw Hill.
- BONNIOL J.J. (1972). — **Les comportements d'estimation dans une tâche d'évaluation d'épreuves scolaires. Etude de quelques-uns de leurs déterminants**. Aix-en-Provence, Université de Provence, thèse de doctorat.
- BONNIOL J.J. (1988). — Entre les deux logiques de l'évaluation, rupture ou continuité ?, **Bulletin de l'ADME**, 88 (3), Craponne.
- BONNIOL J.J., CAVERNI J.P. et NOIZET G. (1972a). — Pour un apprentissage de l'évaluation des tâches scolaires, **Cahiers de Psychologie**, 15, pp. 93-104.
- BONNIOL J.J., CAVERNI J.P., et NOIZET G. (1972b). — Le statut scolaire des élèves comme déterminant de l'évaluation des devoirs qu'ils produisent, **Cahiers de Psychologie**, 15, pp. 83-92.
- BONNIOL J.J., GENTHON M., ROGER M. (1989). — L'évaluation en psychologie, **Bulletin de l'AECSE**, 6, pp. 12-18.
- BOUDON R. (1973). — **L'inégalité des chances**. Paris, Armand Colin, 3<sup>e</sup> édition.
- BOURDIEU P. (1966). — L'école conservatrice - L'inégalité sociale devant l'école et devant la culture, **Revue Française de Sociologie**, 3, pp. 325-347.
- BOURDIEU P. et PASSERON J.C. (1964). — **Les héritiers**. Paris, Ed. de Minuit.
- BOURDIEU P. et PASSERON J.C. (1970). — **La reproduction**. Paris, Ed. de Minuit.
- BRESSOUX P. (1993). — **Les effets des écoles et les classes sur l'apprentissage de la lecture**. Dijon, thèse inédite.
- CARDINET J., TOURNEUR Y. et ALLAL L. (1976). — The symmetry of generalizability theory: application to educational measurement, **Journal of Educational Measurement**, 4 (18), pp. 183-204, et 4 (19), pp. 331-332.
- CARDINET J. (1986a). — **Pour apprécier le travail des élèves**. Bruxelles, De Boeck.
- CARDINET J. (1986b). — **Evaluation scolaire et pratique**. Bruxelles, De Boeck.
- CARDINET J. (1984). — Evaluer sans juger, **Revue Française de Pédagogie**, 88, Paris, INRP.
- CARDINET J. et TOURNEUR Y. (1985). — **Assurer la mesure**. Berne, Peter Lang.
- CARROLL J.B. (1963). — A model of school learning, **Teachers College Record**, 64 (8), 723-733.
- CHARLIER E. (1989). — **Planifier un cours, c'est prendre des décisions**. Bruxelles, De Boeck.
- CHEVALLARD Y. (1985). — **La transposition didactique**. Grenoble, La Pensée Sauvage.
- CHEVALLARD Y. (1990). — Evaluation, véridiction, objectivation, in COLOMB J. — **L'évaluation en révolution**, Paris, INRP, 13-36.
- COLEMAN J.S. et al. (1966). — **Egalité of Educational Opportunity**, Washington, Government Printing Office.
- DAUVISIS M.C. (1988). — La formation à l'évaluation : tendances et questions ouvertes, in GATHERTHURLER M. et PERRENOUD P. (Ed.), **Savoir évaluer pour mieux enseigner, quelle formation des maîtres ?**, **Cahiers des SRS**, 26, 14-17.
- DE KETELE J.M. (1980). — **Observer pour éduquer**. Berne, Peter Lang.
- DE KETELE J.M. (1982). — Le passage de l'enseignement secondaire à l'enseignement supérieur, **Humanités Chrétiennes**, 26 (4), 294-306.
- DE KETELE J.M. (Ed) (1986). — **L'évaluation : approche descriptive ou prescriptive ?** Bruxelles, de Boeck.
- DE KETELE J.M., CHASTRETTE M., CROS D., METTELIN P. et THOMAS J. (1989). — **Le guide du formateur**. Bruxelles, De Boeck.
- DE KETELE J.M. et ROEGIERS X. (1993). — **Méthodologie du recueil d'informations**. Fondements des méthodes d'observation, les questionnaires d'interviews et d'études de documents, Bruxelles, de Boeck.
- DE LANDSHEERE V. (1988). — **Faire réussir - Faire échouer**. Paris, PUF.
- DE LANDSHEERE V. et DE LANDSHEERE G. (1975). — **Définir les objectifs en éducation**. Liège, Georges Thone.
- DE LANDSHEERE G. (1973, 1980) (5<sup>e</sup> éd.). — **Evaluation continue et examens - Précis de docimologie**. Paris, Nathan ; Bruxelles, Ludor.
- DELORME C. (Ed.) (1987). — **L'évaluation en questions**. Paris, ESF.
- DELORME C. et coll. (1986). — **Guide méthodologique pour l'évaluation des PAE**. Grenoble, CRDP.
- DEROUET J.L. (Sid.). — **L'évaluation des établissements scolaires**. AECSE : L'établissement, 139-165.
- DONNAY J. et CHARLIER E. (1991). — **Comprendre des situations de formation**. Bruxelles, De Boeck.
- DOYLE W. (1986). — Paradigmes de recherche sur l'efficacité des enseignants, in CRAHAY M. et LAFONTAINE D. (Ed.). — **L'art et la science de l'enseignement**. Bruxelles, Labor. 435-481.
- DURU-BELLAT M., et MINGAT A. (1985). — De l'orientation en fin de 5<sup>e</sup> au fonctionnement du collège, **Cahiers de l'IREDU**, 42, Dijon.
- DURU-BELLAT M. et MINGAT A. (1988). — Le déroulement de la scolarité au collège : le contexte « fait des différences », **Revue Française de Pédagogie**, 29 (4), 649-666.
- DURU-BELLAT M. et LEROY-AUDOUIN C. (1990). — Les pratiques pédagogiques au CP : structure et incidence sur les acquis des élèves, **Revue Française de Pédagogie**, 93, 5-15.
- FAVRE B. et MONTANDON C. (1989). — **Les parents dans l'école**. Genève, Service de la recherche sociologique. 30.
- FIGARI G. (Ed.) (1988). — **Dispositif et méthodes de**

- formation à la conception et à l'évaluation des PAE.** Grenoble, Université de Grenoble II.
- FIGARI G. (1993). — **Quel référentiel.** Grenoble, Dépt des Sciences de l'Éducation. Manuscrit soumis à publication.
- GIACONIA R. et HEDGES L.V. (1982). — Identifying Features of Effective Open Education, **Review of Educational Research**, 52 (4), 579-602.
- GILLY M. (1980). — **Maître-élève - Rôles institutionnels et représentations.** Paris, PUF.
- HADJI C. (1992). — **L'évaluation des actions éducatives.** Paris, PUF.
- HAMELINE D. (1980). — **Les objectifs pédagogiques en formation initiale et en formation continue.** Paris, ESF (2<sup>e</sup> éd.).
- HAINAUT L. (d') (1980a). — **Des fins aux objectifs de l'éducation,** Paris-Bruxelles, Nathan-Labor.
- HAINAUT L. (d') (1980b). — **La régulation dans les systèmes éducatifs.** Guide méthodologique, UNESCO, 33.
- KAUFMAN J.P. (1976). — La perception des élèves par des enseignants : comparaison des résultats fournis par deux méthodes d'analyse, **Bulletin de Psychologie**, 326 (30), 30-45.
- JENCKS C. et al. (1972). — **Inequality : A Reassessment of Family and Schooling in America.** New York, Booi Books.
- LA GARANDERIE A. (de) (1982). — **Pédagogie des moyens d'apprendre - Les enseignants face aux profils pédagogiques.** Paris, Centurion.
- LAMONTAGNE C. (1984). — **Vers une pratique du profil d'apprentissage.** Saint-Hubert (Qc), Institut de recherche sur le profil d'apprentissage.
- LAMONTAGNE C. (1987). **Test pour détermination du style d'apprentissage.** Saint-Hubert (Qc), Institut de recherche sur le profil d'apprentissage.
- LONDEIX H. (1982). — Structure fonctionnelle de l'opinion des professeurs sur leurs élèves, **L'orientation scolaire et professionnelle**, 11, 169-190.
- MEIRIEU P. (1985). — **L'école, mode d'emploi.** Paris, ESF.
- MEIRIEU P. (1987). — **Apprendre, oui, mais comment ?** Paris, ESF.
- MEURET D. (1990). — Les méthodes d'analyse des établissements scolaires et leur usage pour l'évaluation, **Éducation et Formation**, 22, 3-5.
- MINGAT A. (1983). — Évaluation analytique d'une action Zone d'éducation prioritaire en cours, Préparation, **Cahiers de l'IREDU**, 37, Dijon.
- MINGAT A. (1984). — Les acquisitions scolaires au CP : les origines des différences ? **Revue Française de Pédagogie**, 69, 49-62.
- MINGAT A. (1987). — Sur la dynamique des acquisitions à l'école élémentaire, **Revue Française de Pédagogie**, 79, 5-14.
- MINGAT A. (1991). — Expliquer la variété des acquisitions au cours préparatoire : les rôles de l'enfant, la famille et l'école, **Revue Française La Pédagogie**, 95, 47-63.
- MOLLO S. (1970). — **L'école dans la société.** Paris, Dunod.
- MONTANDON C. (1991). — **L'école dans la vie des familles.** Genève, Service de la recherche sociologique, 32, 3.
- MONTEIL J.M. (1985). — **Dynamique sociale et synthèse de formation.** Paris, UNMFREO.
- MONTEIL J.M. (1989). — **Éduquer et former - Perspectives psychosociales.** Grenoble, Presses Universitaires de Grenoble.
- MORIN E. (1977). — **La méthode.** Paris, Seuil.
- NOIZET G. et CAVERNI J.P. (1978). — **Psychologie de l'évaluation scolaire.** Paris, PUF.
- NUNZIATI G. (1990). — Pour construire un dispositif d'évaluation formatrice, **Cahiers Pédagogiques**, 280.
- PERRETI A. (de) (1991). — **Organiser des formations.** Paris, Hachette.
- PERRETI A. (de) et al. (1981). — **Recueil d'instruments et de processus d'évaluation formative.** Paris, INRP.
- PERRETI A. (de) (1985). — Esquisse d'un fondement théorique de la pédagogie différenciée, **Les Amis de Sèvres**, 1, Sèvres, CIEP.
- PERRENOUD P. (1979). — Des différences culturelles aux inégalités scolaires : l'évaluation et la mesure dans un enseignement différencié, in ALLAL L., CARDINET J. et PERRENOUD P. — **L'évaluation formative dans un enseignement différencié.** Berne, Peter Lang.
- PERRENOUD P. (1984). — **La fabrication de l'excellence scolaire.** Genève-Paris, Librairie Droz.
- PERRENOUD P. (1991). — Pour une approche pragmatique de l'évaluation formative, **Mesure et évaluation en éducation**, 4 (13), 49-81.
- PIÉRON H. (1963). — **Examens et docimologie.** Paris, PUF.
- SCALLON G. (1981). — **La construction d'un test diagnostique selon des facettes.** Québec, Université de Laval (monopolisée en mesure et évaluation).
- SCALLON G. (1986). — **L'évaluation formative des apprentissages.** Québec, Les Presses de l'Université Laval.
- SCRIVEN M. (1967). — The methodology of evaluation, in **Perspectives of curriculum evaluation.** Chicago, Rand Mac Nally, 39-83.
- SCRIVEN M. (1983). — Evaluation Ideologies, in MADAUS G.F., SCRIVEN M. et STUFFLEBEAM D.L. — **Evaluation Models.** Boston, Kluwer Nijhoff.
- STAKE R.G. (1967). — The countenance of educational Evaluation, **Teachers College Record**, 68, 523-540.
- STAKE R.E., (1975). — **Evaluating the Arts in Education a responsive Approach.** Columbus, Merrill.
- STAKE R.E. (1976). — A Theoretical Statement of responsive Evaluation, **Studies in Educational Evaluation**, 2, 19-22.
- STUFFLEBEAM S.L. et al. (1980). — **L'évaluation en éducation et la prise de décision.** Ottawa, NHP.
- STUFFLEBEAM S.L. et SHINKFIELD A.J. (1985). — **Systematic Evaluation.** Boston, Kluwen et Nijhoff Publishing.
- WEISS J. (1986). — La subjectivité blanchie ? in DE KETELE J.M. (Ed.). — **L'évaluation : approche descriptive ou prescriptive.** Bruxelles, De Boeck.