

# L'OBJECTIVITÉ DE L'ÉVALUATION

Jean CARDINET

RECHERCHES

87.104 – Juin 1987

# L'OBJECTIVITÉ DE L'ÉVALUATION

Jean CARDINET

IRDP

INSTITUT ROMAND DE RECHERCHES ET DE DOCUMENTATION PÉDAGOGIQUES  
Faubourg de l'Hôpital 43 CH 2000 Neuchâtel Tél. (038) 24 41 91

CARDINET, Jean. - L'objectivité de l'évaluation /  
Jean Cardinet, - Neuchâtel : Institut romand de recherches et de documentation pédagogiques, 1987. - 12 p. ;  
30 cm. - (Recherches ; 87.104)  
Bibliogr. p. 11-12

Evaluation  
Objectivité  
Evaluation formative  
Recherche en éducation

*La reproduction, totale ou partielle, des publications de l'IRDP est en principe autorisée, à condition que leur(s) auteur(s) en aient été informé(s) au préalable et que les références soient mentionnées.*

## L'OBJECTIVITE DE L'EVALUATION

---

Jean CARDINET

### Résumé

Il semble équitable que les examens traitent tous les élèves de la même façon. Mais il faudrait aussi que les questions posées révèlent ce que les élèves ont appris. Or ces deux exigences sont incompatibles, parce que les maîtres n'enseignent pas de la même façon. De plus, les élèves peuvent réussir, ou échouer, selon la façon dont ils sont interrogés. Choisir une forme d'examen standardisée n'améliorerait pas la précision de la mesure. Il faut préférer une évaluation formative, effectuée dans le cadre de la classe. Pour des bilans extérieurs, présenter directement les productions des élèves évite plusieurs sources d'erreurs importantes.

## DIE OBJEKTIVITÄT DER BEWERTUNG

---

Jean CARDINET

### Zusammenfassung

Es erscheint recht und billig, dass die Prüfung alle Schüler gleich behandeln. Allerdings müssten die gestellten Aufgaben aber auch das aufzeigen, was die Schüler gelernt haben. Nun sind aber diese zwei Ansprüche unvereinbar, weil die Lehrer nicht im selben Stil unterrichten. Dazu kommt noch, dass die Schüler Erfolg oder Misserfolg ernten können, je nach der Art und Weise, in der sie fragt werden. Die Wahl einer standardisierten Examensform würde die Genauigkeit der Bewertung nicht verbessern. Eine formative Evaluation ist vorzuziehen, die im Rahmen der Klasse vorgenommen wird. Für die Bilanz nach aussen kann man sagen, dass die direkte Präsentation der Schülerarbeiten viele ausschlaggebende Fehlerquellen vermeiden kann.

# L'OBJECTIVITE DE L'EVALUATION

J. CARDINET

## 1. L'OBJECTIVITE EN QUESTION

La critique des examens n'est pas nouvelle. Déjà la recherche de Laugier et Weinberg, publiée en 1938, mettait en cause la fidélité des correcteurs du baccalauréat, et les recherches de docimologie de Piéron (1969) sont venues confirmer les doutes concernant l'objectivité de telles épreuves. Mais une série d'approches nouvelles, surtout sociologiques et psycho-sociologiques, ont pu récemment éclairer le problème sous des angles différents, dont il est important de tenir compte.

De Landsheere (1979) définit l'objectivité comme "le caractère de ce qui donne une image non déformée... des choses". L'objectivité d'un test est pour cet auteur "le fait qu'il est relativement exempt d'erreurs de jugement ou de correction ... que ses résultats dépendent seulement de la performance du sujet".

Si, comme le dit encore De Landsheere, "le garant de cette objectivité réside dans l'emploi de règles de correction et de notation précises", on peut se demander pourquoi, après un demi-siècle de recherches, le système scolaire n'est pas parvenu à assurer cette objectivité dans le jugement des performances des élèves.

Formuler ainsi la question est tendancieux, répondra-t-on, parce que c'est jouer sur deux sens différents du mot "objectivité", pris chez De Landsheere comme une des qualités métrologiques d'un test, et non pas comme valeur de vérité d'un jugement. Mais ces deux sens sont adjacents et la confusion n'est pas fortuite. Malgré ce qu'en dit un certain discours officiel, l'évaluation scolaire a beaucoup de raisons de rester subjective; le démontrer sera l'objet des pages qui suivent.

## 2. L'OBJECTIVITE NECESSAIRE

Pour faire fonctionner le système scolaire, et notamment pour assurer la continuité de l'enseignement lorsque les élèves passent d'une classe à une autre, des points de repère sont nécessaires : des programmes d'une part, des bilans d'apprentissage par rapport à ces programmes, d'autre part.

\* Cette conférence sera présentée lors du IIe CONGRES MONDIAL BASQUE, à Bilbao (Congrès d'éducation, Aire III: Psychopédagogie et Didactique), en octobre 1987.

Il semble aller de soi que chacun des partenaires de l'école aurait intérêt à savoir exactement ce qui a été appris. Ce serait l'intérêt des élèves de manifester quelles sont les acquisitions sur lesquelles leurs nouveaux apprentissages vont pouvoir s'appuyer. Ce serait aussi, pour les enseignants, un retour d'information essentiel sur l'efficacité de leur travail.

Du point de vue de l'équité, enfin, cela paraît même un devoir pour tous ceux qui doivent prendre des décisions concernant la carrière scolaire des enfants de suivre des règles uniformes, s'appuyant sur des évaluations objectives de leur niveau réel de connaissance.

Mager (1984) donne dans ce but de précieux conseils sur la façon dont un maître peut mesurer correctement les résultats de son enseignement. Il faut pour cela que celui-ci définisse précisément les objectifs éducatifs qu'il vise, puis qu'il choisisse des performances observables correspondant à ces objectifs, enfin qu'il maintienne, pour l'opération de mesure, les mêmes conditions de travail que celles qu'il avait prévues en définissant les objectifs. Tout au long de son ouvrage, Mager insiste sur la correspondance qu'il faut assurer entre ce qui est enseigné et ce qui est contrôlé: c'est une condition essentielle pour que la mesure donne "une image non déformée" des résultats obtenus.

Cependant l'évaluation ne saurait se limiter à ce contrôle. Mager le précise d'ailleurs lui-même (traduction p.8): "L'acte d'évaluation compare une mesure avec un standard et porte un jugement sur la comparaison". Comme Ardoino (1976) l'a, le premier, mis en évidence, l'évaluation fait ainsi référence à des valeurs extérieures au système, alors que le contrôle pourrait être confié à un automate de mesure à l'intérieur du système. Il est symptomatique, par exemple, que Mager ne fasse nulle part allusion à la détermination du seuil de suffisance acceptable: ce serait dépasser le contrôle, et par là même ouvrir la porte à la subjectivité. On peut donc déjà pressentir les limites de cette objectivité que l'on dit nécessaire.

### 3. L'OBJECTIVITE SOUHAITEE

Les recherches effectuées récemment sur le fonctionnement du système scolaire viennent jeter une lumière crue sur le manque d'objectivité effectif de l'évaluation des élèves. Un exemple sera présenté pour chacun des trois pays francophones d'Europe.

Duru et Mingat (1987) ont étudié les pratiques actuelles des conseils de classe qui décident en France du redoublement, ou de l'orientation scolaire ultérieure, des élèves en fin de cinquième. Ils ont cherché ensuite ce qui adviendrait si ces conseils appliquaient uniformément les règles qui décrivent le mieux leur stratégie de décision d'ensemble. (L'écart entre les décisions effectives et les décisions ainsi simulées représente donc l'effet du manque d'accord d'une classe à l'autre sur les règles à suivre pour l'orientation). Ils observent alors que "153 élèves, actuellement redoublants, (soit le 47% de la population des redoublants étudiée) seraient admis dans la classe supérieure; simultanément, 251 parmi les élèves actuellement autorisés à passer en quatrième (soit le 18,5% des élèves actuellement admis) se verraient refuser cette orientation". Cette influence énorme de facteurs indépendants de la performance du sujet est d'autant plus inquiétante qu'on voit par ailleurs que les élèves faibles, admis néanmoins en quatrième, y suivent en moyenne une scolarité normale: les critères de la décision de redoublement, même s'ils

étaient appliqués uniformément, auraient donc eux aussi peu de validité. Peut-on dire que l'évaluation des élèves en France est "relativement exempte d'erreurs de jugement" ?

La situation n'est pas meilleure en Belgique. Grisay (1984, 1986) y a étudié les pratiques d'évaluation-bilan dans 52 écoles primaires. L'auteur conclut que la décision de redoublement tient de la loterie: 5% seulement de sa variance peut être prédite par une épreuve de français (qui avait été validée par ailleurs). Même les notes des instituteurs n'expliquent que 20% de cette variance. En cherchant, comme Duru et Mingat, à simuler au mieux la stratégie de décision utilisée par les maîtres, l'auteur reste incapable de comprendre la justification du redoublement dans à peu près un cas sur deux. Des critères autres que le rendement de l'élève doivent intervenir. Un questionnaire suggère les facteurs suivants: pressions différentes des parents, crainte de perdre des élèves, de gêner un dédoublement de classe, réputation de sélectivité de l'école et demandes des collègues suivants. De toute façon, les enseignants n'ont qu'une idée très personnelle des exigences de chaque degré. Dans une classe, l'examen de fin de deuxième année en français consistera à mettre au pluriel une phrase de quatre mots simples, dans une autre, une liste de vingt mots, dont dix-sept sont des difficultés orthographiques. Sur 650 questions de français données aux examens de deuxième année, 450 sont des exercices de grammaire ou d'orthographe grammaticale, alors que la grammaire formelle n'apparaît officiellement au programme qu'au degré suivant. Il en résulte de très grandes variations d'exigences entre classes, dont les enseignants n'ont pas conscience. Comme ils distribuent leurs notes de la même façon partout, indépendamment du niveau effectif de leur classe, ils sont amenés involontairement à surestimer en général leurs meilleurs élèves et à sous-estimer les plus faibles. On voit ainsi comment se "fabrique" l'excellence scolaire dont Perrenoud (1984) a étudié la différenciation progressive au sein des classes primaires.

Des constats aussi inquiétants pour l'objectivité des notes ont été faits en Suisse. Flammer et Perrig-Chiello (1979) ont comparé les notes données dans divers cantons. Pour une même branche, la distribution peut être de près d'un sigma plus élevée dans une région que dans une autre. D'ailleurs, les taux de sélection pour des études supérieures varient tout autant, le pourcentage d'admissibles à l'université étant actuellement de 5% d'une tranche d'âge en Suisse centrale, contre 22% à Bâle ou 26% à Genève. Ces différences ne peuvent pas tenir à la capacité personnelle des élèves. Une étude détaillée des notes données au cours de la scolarité secondaire (Pfister, 1975) a montré une baisse constante des notes, au fur et à mesure que les élèves progressaient dans leurs études. Il en résulte que certaines classes expérimentales qui commencent l'apprentissage d'une seconde langue avant les autres et sont donc en avance sur eux, reçoivent des notes inférieures et sont ainsi pénalisées pour leur supériorité.

Est-il nécessaire de souligner la gravité sociale de ces inégalités devant l'évaluation? Les travaux récents de psychologie sociale expérimentale (Monteil, 1987), qui démontrent que des notes données au hasard déterminent chez les élèves des performances ultérieures en accord avec ces évaluations, ne peuvent que renforcer les craintes de tous ceux qui s'intéressent à l'éducation. Pourquoi n'améliore-t-on pas les procédures d'évaluation, en suivant par exemple la démarche proposée par Mager ?



#### 4. L'OBJECTIVITE IMPOSSIBLE

Le défi vaudrait la peine d'être relevé, mais les difficultés à prévoir doivent pour cela être connues. Certaines sont présentées ci-dessous, pour chacune des étapes proposées par cet auteur.

##### 4.1. Objectifs différents

L'essentiel, pour Mager, est de tester ce que l'on a enseigné. Mais comment alors préparer des examens comparables, si les maîtres n'enseignent pas la même chose ? Or c'est manifestement ce qui se passe.

Du point de vue des contenus, les programmes devraient assurer l'homogénéité désirée, mais on sait que les maîtres n'assimilent eux-mêmes les notions nouvelles que progressivement, et reviennent souvent à ce qu'ils maîtrisaient le mieux auparavant. Une enquête de Tourneur (1981) a même révélé que des nouveaux programmes, en Belgique, n'avaient été distribués à tous les enseignants que plusieurs années après leur mise en application. Il est donc certain que les programmes sont couverts très inégalement dans les classes.

C'est souvent d'ailleurs en raison d'interprétations différentes des directives reçues que les enseignants se distinguent. Les uns privilégient en mathématique l'acquisition des algorithmes, les autres les démarches de recherche et de raisonnement; d'autres encore se centrent sur les applications possibles. Tout cela est au programme, mais le poids accordé à ces différents apprentissages peut varier très largement.

Le choix de la progression, généralement laissé aux instituteurs, conduit à un découpage très différent des connaissances. Ainsi, l'emploi de différentes méthodes de lecture ne permet pas de comparer les acquisitions des élèves avant que ces derniers n'aient fini d'apprendre à lire.

##### 4.2. Performances différentes

La façon dont l'élève doit manifester que l'objectif est atteint peut varier d'une classe à l'autre, notamment en raison de la conception que le maître se fait du résultat visé, de la démarche requise et de la capacité mesurée.

Le résultat peut être observé à court, ou à plus long terme. Dans le premier cas, le maître contrôlera des savoirs limités, de façon analytique; dans le second, il vérifiera que l'élève est capable d'utiliser ses connaissances pour résoudre un problème plus global, demandant par exemple une recherche d'information au lieu d'une simple lecture orale. Le contrôle d'objectifs terminaux d'intégration, suggéré par De Ketele (1984) rend la performance plus significative, mais en même temps plus aléatoire, dans la mesure où un transfert est demandé à une situation nouvelle, et cette conception du résultat à évaluer est loin d'être généralisée.

Le démarche employée pour résoudre le problème posé peut, ou non, être prise en compte dans le contrôle. Certains maîtres se satisfont de l'exactitude d'un calcul, d'autres mettent en cause la signification d'un résultat obtenu par tâtonnement, par exemple (Brun, 1979). Certains veulent, au laboratoire, contrôler un savoir-faire, d'autres un savoir-dire. Lorsque l'élève apprend

une langue, doit-il pouvoir formuler les règles qu'il applique, ou simplement s'exprimer correctement ? Les avis divergent.

La capacité mesurée varie selon la performance choisie par le maître. De la simple question de mémorisation à la résolution d'un problème nouveau, ou même à l'analyse de relations abstraites, le niveau de l'objectif testé s'élève, et en même temps la part d'intelligence générale requise. Des types d'élèves différents seront valorisés par chaque genre d'épreuve.

#### 4.3. Conditions d'observation différentes

Le résultat du contrôle dépend de la façon dont il est mené, et d'abord du moment où il est introduit: effectué trop tôt, le bilan n'a pas de sens; trop longtemps après l'apprentissage, c'est la résistance à l'oubli que l'on mesure. Comment arriver au bon moment pour tous les élèves ?

La forme du contrôle, surtout, a donné lieu à de nombreuses recherches qui, toutes, montrent la relativité des évaluations possibles. En mathématique, Vergnaud (1981) a mis en évidence l'influence déterminante de la structure logique du problème, pour une notion mathématique donnée, comme celle de la soustraction. Selon que l'on interroge l'élève sur le résultat de l'opération, sur l'état initial, ou sur la transformation apportée, la réussite peut être retardée de plusieurs années. L'habillage du problème (selon que l'on parle de mètres ou de francs, d'un contexte familier ou de réalités lointaines pour l'enfant, etc.) contribue beaucoup aux variations de difficulté. La formulation verbale de la question, enfin, joue un rôle tellement prépondérant que ces problèmes verbaux mesurent sans doute plus la compréhension de lecture que les connaissances en mathématique (Cardinet, 1986). Que contrôle-t-on alors ?

Le mode de correction ajoute sa part de variance à la mesure du résultat. Ce fut même le premier facteur perturbateur mis en évidence par les études docimologiques. La psychologie du notateur et les principaux effets qui modifient son jugement ont été réétudiés plus récemment (Noizet et Caverni, 1978; Bonniol, 1981), mettant en évidence les sources du manque de fidélité découvert précédemment (effets d'ancrage, de contraste, etc.), sans résoudre pourtant les problèmes pratiques correspondants. Faudrait-il ramener tous les examens à une forme unique, celle des questions à choix multiples ? Ce serait oublier que cette forme d'interrogation introduit ses propres biais: pour beaucoup d'élèves, il est plus facile de reconnaître la bonne réponse que de la formuler entièrement (Cardinet et Dubosson, 1969).

#### 4.4. Exigences des évaluateurs différentes

Dans le cadre de la classe, l'appréciation du résultat se traduit, traditionnellement, par la mise d'une note. Or aucune fonction objective ne relie le nombre, ou la qualité, des réponses avec cette appréciation. Chevallard (1986) a bien démontré l'arbitraire du jugement ainsi porté, qui ressemble plus à un marchandage avec la classe qu'à une mesure. Selon le moment où elle se situe, l'interrogation servira à obliger les élèves à travailler davantage, ou au contraire à remonter une moyenne de classe compromettante pour l'enseignant.

Si pour un même évaluateur les exigences évoluent, il va de soi qu'elles varient d'un enseignant à l'autre, comme le montre Boumard (1978) en analysant un conseil de classe. Elles varient surtout selon les écoles d'après les recherches de Duru et Mingat (1987), en raison de cultures locales spécifiques.

Elles varient enfin selon le niveau socio-culturel du milieu social d'origine des élèves, des effets d'ancrage déplaçant les échelles d'évaluation pour les adapter à la performance moyenne observée par l'enseignant.

#### 4.5. Significations de l'examen différentes

Une dernière source d'hétérogénéité a été mise en évidence récemment par les travaux de psychologie sociale de l'éducation. Perret-Clermont (1979) a montré qu'un enfant passait rapidement d'un stade de développement opératoire au stade supérieur dans un contexte d'examen où la dynamique du groupe social favorisait sa réflexion, par opposition au contexte de l'examen classique, mené par un adulte. Était-ce fausser les conditions d'examen, comme auraient tendance à la dire les piagétiens orthodoxes ? C'était plutôt montrer que toute situation d'évaluation, quelle qu'elle soit, affecte inévitablement la probabilité de réussite, et qu'il n'existe pas de situation de référence permettant une mesure objective. Schubauer-Léoni et Perret-Clermont (1985) ont pu observer les mêmes phénomènes à l'oeuvre lors d'interrogations relatives à la mathématique: les enfants donnent un sens à la question qui leur est posée d'après ce qu'ils comprennent de la situation d'examen dans son ensemble. Carraher (1987) a montré également que des enfants qui avaient résolu correctement des problèmes mathématiques dans une situation réelle se révélaient incapables de le faire à nouveau en situation de classe: ils ne mobilisaient plus les mêmes démarches de pensée et perdaient alors la référence à la réalité qui les avait guidés dans le premier cas. Laquelle est la compétence véritable de ces élèves ?

Il n'est pas plus facile de répondre pour d'autres situations où l'examen, apparemment objectif, parce que le même pour tous, recèle en réalité un biais par rapport à un autre point de référence. Les épreuves de sélection scolaire qui sont encore passées dans certains cantons suisses à l'âge de 11 ans favorisent massivement les filles, qui ont atteint plus tôt que les garçons certains traits de maturité intellectuelle et affective. La différence serait sans doute moins grande quatre ans plus tard. La seconde mesure serait-elle alors plus "objective" que la première ? Le Tribunal Fédéral n'a pas voulu l'admettre, sentant bien que chercher une référence ultime aurait mené à des discussions sans fin.

#### 4.6. Que répondre à Mager ?

Les propositions de cet auteur sont parfaitement judicieuses si, comme il le demande, c'est chaque enseignant pris individuellement qui définit les objectifs, performances et conditions par rapport auxquels il jugera de la réussite de l'apprentissage de ses élèves. L'erreur, par contre, serait de croire que la "mise en ordre" des procédures d'évaluation classe par classe assurerait l'objectivité des examens à l'intérieur du système scolaire dans son ensemble. L'arbitraire des choix nécessaires concernant chacune des étapes ci-dessus laisse penser que le résultat de ces démarches "raisonnées" ne serait pas moins hétérogène que celui des démarches "intuitives" actuelles. Les résultats des évaluations scolaires sont "sans commune mesure les uns avec les autres", parce que les élèves eux-mêmes ne sont jamais dans des conditions comparables.

Bien qu'elle paraisse négative, cette conclusion n'en est pas moins importante. Elle conduit à abandonner l'idée si répandue qu'il existerait une situation d'examen "de référence", un test privilégié, qui permettrait de savoir ce que les élèves savent vraiment, un mètre étalon pour trait latent, satisfaisant au modèle de Rasch, par exemple. Elle conduit au contraire à

s'apercevoir que la performance des élèves est une fonction de plusieurs variables, dans laquelle les "conditions d'examen" rendent compte d'une part de variance importante.

## 5. L'OBJECTIVITE REJETEE

Cette conclusion peut paraître hâtive, et trop peu audacieuse. Les révolutionnaires français n'ont-ils pas créé le mètre, pour mettre de l'ordre dans un ensemble de mesures également disparates ? Mais il suffit de reprendre les sources d'hétérogénéité mentionnées au point précédent pour voir ce qu'impliquerait la création d'une échelle de référence unique.

Il faudrait d'abord définir exhaustivement les objectifs visés dans l'enseignement, de façon beaucoup plus précise que pour les programmes ou objectifs officiels d'aujourd'hui. Il faudrait spécifier tous les comportements terminaux à atteindre et le chemin pour y parvenir. Ceci impliquerait de prévoir toutes les activités à faire faire aux élèves, avec leur ordre de présentation et le rythme d'avancement à respecter, faute de quoi le contenu et le moment des contrôles serait mal adapté.

Il faudrait ensuite décider de performances intermédiaires, c'est-à-dire d'objectifs d'intégration jalonnant la progression précédente, en précisant quels aspects de la démarche seraient obligatoires, quelles métaconnaissances seraient requises, quel degré de transfert serait demandé aux élèves.

Il faudrait surtout constituer pour chaque objectif des moyens d'observation dont on aurait vérifié la comparabilité. Enfin, toutes les classes et tous les milieux devraient être jugés sur les mêmes bases, non seulement formellement, mais aussi du point de vue des intéressés.

En admettant même que les énormes travaux préparatoires impliqués puissent être réalisés, les conséquences sociales d'une telle entreprise seraient inacceptables, aussi bien pour les enseignants que pour les parents: seuls des ordinateurs pourraient donner un enseignement aussi standardisé; et comment s'adaptent-ils aux individualités des élèves ?

Beaucoup des sources de différences mentionnées plus haut sont en fait des adaptations fonctionnelles à une réalité imprédictible: la forme particulière que prend la progression d'une classe en interaction avec un enseignant. Le niveau d'objectif visé doit être variable, pour que le maître puisse rencontrer les élèves où ils sont et faire cheminer la classe à un rythme quelle puisse suivre. Le manque de standardisation d'un examen oral est aussi la condition nécessaire pour que s'engage un dialogue, permettant au maître de comprendre les représentations de l'élève interrogé. L'ancrage de l'échelle d'appréciation de l'enseignant sur la moyenne de sa classe est sa façon de différencier au mieux la perception qu'il a de ses élèves. Le flou des échelles d'appréciation est, comme le dit Perrenoud "le jeu dans l'engrenage pour éviter que la machine ne se bloque", lorsque paraissent socialement inacceptables aussi bien les inégalités, que les mesures égalitaires qui les réduiraient.

Le coût de l'harmonisation paraît ainsi exorbitant en regard des bénéfices qu'elle apporterait dans la situation actuelle.

## 6. DES BASES POSSIBLES POUR L'EVALUATION

Si l'objectivité totale est hors d'atteinte, cela n'implique pas de renoncer à améliorer la situation existante. Des instruments de mesure peuvent être utiles aux enseignants, si ces derniers prennent soin, comme les psychologues ou les médecins, de les interpréter cliniquement, c'est-à-dire par rapport au contexte global où se situe chaque élève.

### 6.1. Evaluation formative

La première possibilité de progrès, et la plus facile à réaliser, est de centrer l'évaluation scolaire sur une démarche formative. Le manque de point de repère pour décider si le niveau d'un élève est suffisant ou non n'est guère gênant pour un enseignant, tant que ce dernier n'est pas obligé de mettre une note à une copie. Ce maître peut, par contre, corriger le travail et donner un retour d'information détaillé à l'élève, l'aidant ainsi à progresser, quel que soit son niveau. Insérer des boucles de rétroaction dans la démarche didactique est toujours possible. Communiquer aux élèves, à cette occasion, la nature exacte des objectifs et spécifier les critères de la performance que l'enseignant considère comme acceptable (selon les propositions de Mager), voilà le premier et le plus grand pas à faire en direction d'une évaluation scolaire satisfaisante, qui apporte des informations à l'élève, et non sur l'élève.

### 6.2. Evaluation d'étape

Dans le cadre de chaque classe, il est possible de choisir, parmi l'ensemble des objectifs souhaitables et des performances observables, un certain nombre de savoirs et de savoir-faire qui feront l'objet du contrat didactique. Le maître et les élèves doivent pouvoir communiquer à son sujet, à l'aide de critères explicites. La maîtrise d'une étape est acquise quand ces critères sont satisfaits. Cette façon de faire n'exige pas la comparabilité des progressions d'une classe à une autre, mais ne permet pas non plus la comparabilité des résultats. Elle constitue simplement un complément possible et souhaitable à l'évaluation formative précédente, qui donne aux partenaires, au sein de la classe, des points de repère relativement objectifs pour communiquer entre eux. Les propositions de Mager peuvent conduire les enseignants jusqu'à ce niveau d'objectivité.

### 6.3. Echelles de référence

Les maîtres voudraient bien, cependant, pouvoir aller un peu plus loin encore, en situant leur classe par rapport aux autres. Sont-ils en avance, ou en retard ? Ont-ils satisfait leur contrat en fin d'année ? Une réponse exacte à ces questions paraît impossible, du fait des différences entre les enseignements, déjà mentionnées plus haut. Dans certains cas, pourtant, une réponse inexacte serait peut-être préférable à l'absence totale d'information.

C'est dans ce sens qu'on pourrait recommander aux enseignants de constituer des épreuves correspondant au "tronc commun" de ce qu'ils ont présenté dans leur classe et de les utiliser pour comparer leurs résultats à la moyenne des autres. (Des moyens existent pour assurer la confidentialité de ces statistiques). Pour établir des points de repères plus précis, des banques d'items étalonnés peuvent servir à constituer des épreuves de référence adaptées à la

progression suivie dans chaque classe. Il faudra toujours tenir compte, cependant, du risque que les populations, les moments, ou les conditions d'examen, rendent la comparaison erronée.

#### 6.4. Portefeuille de production

Si la constitution d'échelles de référence est si délicate, pourquoi ne pas s'en passer ? Dans la classe même, les évaluations formatives et d'étape peuvent suffire. C'est pour communiquer avec l'extérieur (parents, directeur d'école, etc.) que des rapports paraissent nécessaires. L'idée commence à se répandre de répondre à ces demandes d'information en présentant directement les productions des élèves. L'enseignant qui a sous les yeux une rédaction est bien mieux renseigné sur la compétence de son auteur que par une note ou une appréciation. Pourquoi n'en serait-il pas de même pour un inspecteur, un futur employeur, etc. ? On évite ainsi deux sources d'erreur majeures : celle du codage du premier correcteur, et celle du décodage de celui qui doit interpréter le rapport. On évite en même temps de perdre l'information concernant tous les aspects qualitatifs du travail, auxquels les destinataires peuvent s'intéresser de points de vue très différents et qu'il est donc impossible de résumer.

En conclusion, on améliorera l'évaluation en essayant de répondre le plus directement possible aux besoins d'information, lorsqu'ils s'expriment dans la classe et hors de la classe. On évitera les solutions bureaucratiques, qui cachent sous une apparence de quantification trompeuse leur manque d'objectivité foncier. On s'inspirera des expériences de l'éducation des adultes qui, par le biais de la reconnaissance des acquis expérimentiels, ouvrent des voies nouvelles à l'évaluation-bilan.

### 7. LA STANDARDISATION ILLUSOIRE

Même dans le domaine des sciences physico-chimiques, la standardisation des contrôles pour assurer des mesures comparables est une opération contre-nature. On doit consacrer de gros efforts à amortir les vibrations, supprimer les poussières, contrôler la température, etc.

Dans les conditions habituelles des écoles, des précautions beaucoup plus nombreuses encore seraient nécessaires pour contrôler les conditions d'observation, si l'on voulait rendre comparables les mesures d'apprentissage envisagées. On serait amené à choisir l'une ou l'autre des deux solutions classiques en matière de contrôle : la fixation des conditions, ou leur échantillonnage aléatoire. Dans les deux cas, le résultat serait vide de sens. Si les conditions étaient fixées, on continuerait à ne rien savoir de la performance des élèves dans une infinité d'autres situations tout aussi pertinentes. Si les conditions étaient choisies au hasard dans un univers de problèmes assez large, la connaissance de la tendance centrale et de la dispersion des résultats d'un élève laisserait subsister une telle indétermination quant à ses capacités, qu'on voit mal quel usage pratique on pourrait faire de cette information purement statistique.

Peut-être parviendra-t-on un jour à analyser les sources de difficulté qui interviennent dans chaque tâche scolaire et à affiner la prédiction de succès ou d'échec pour des univers de généralisation choisis à volonté ; mais on est encore très loin de ce niveau de connaissance.

Même si c'était le cas, on n'aurait réussi encore qu'à améliorer le contrôle (au sens d'Ardoïno), alors que le problème de l'évaluation, c'est-à-dire du sens à donner au résultat observé, demeurerait entier. Un exemple simple suffit à le faire comprendre.

Lorsqu'on fait l'expérience d'appliquer deux fois la même épreuve scolaire à quelques années de distance, on s'aperçoit toujours, au moment d'interpréter les résultats, que tous les éléments du contexte ont changé de sens : les objectifs ont évolué, l'enseignement s'est transformé, les élèves sont différents. Même si les questions sont restées formellement les mêmes, elles ne sont plus comprises actuellement de la même façon qu'autrefois. Aussi "objectives" que soient les différences de moyennes, le sens à leur donner ne peut pas automatiquement faire l'accord des observateurs. C'est pourquoi, dans un champ aussi complexe que celui de l'éducation, les interprétations ne pourraient rester objectives et complètement contrôlables que si l'on renonçait volontairement à tenir compte de toute une part de la réalité.

S'il en est ainsi, on ne pourra jamais soumettre les comportements d'apprentissage, aussi bien programmés soient-ils, à une interprétation automatique, ni les réduire au seul constat de réussite ou d'échec. L'évaluation déborde le contrôle. La signification de la situation où s'inscrit l'examen modifiera toujours le sens et la portée de chacune des questions que l'on avait posées.

\* \* \*

Références

Ardoino, J. Préface du livre de M. Morin, L'imaginaire dans l'éducation permanente : analyse du discours des formateurs. Paris : Bordas/Gauthier-Villars, 1976.

Bonniol, J.-J. Déterminants et mécanismes des comportements d'évaluation d'épreuves scolaires. Bordeaux: Thèse, Université de Bordeaux II, 1981.

Boumard, P. Un conseil de classe très ordinaire. Paris: Stock, 1978.

Brun, J. L'évaluation formative dans un enseignement différencié de mathématique. In: L. Allal, J. Cardinet et Ph. Perrenoud (Ed.). L'évaluation formative dans un enseignement différencié. Berne: Peter Lang, 1979, p. 170-181.

Cardinet, J. A la recherche d'une évaluation par objectifs. In: J. Cardinet, Evaluation scolaire et pratique. Bruxelles: De Boeck, 1986, p. 11-95.

Cardinet, J. et Dubosson, J. Comparaison des propriétés métriques de tests à réponse libre et à choix multiple. Revue suisse de psychologie pure et appliquée, vol. 28, N° 1 (1969), p. 12-27.

Carraher, T.N. Mathematics as a personal and social activity. Communication présentée au Congrès international "Le fonctionnement de l'enfant à l'école", Université de Poitiers, 17-20 juin 1987.

Chevallard, Y. Vers une analyse didactique des faits d'évaluation. In: J.-M. De Ketele, L'évaluation: approche descriptive ou prescriptive ? Bruxelles: De Boeck, 1986, p. 31-59.

Duru, M. et Mingat, A. Le redoublement au Collège, en France : éléments pour une évaluation. Dijon: IREDU, 1987.

Flammer, A. et Perrig-Chiello, P. Die Not mit der Schulnote : Was wollen wir eigentlich ? Education et Recherche, vol. 1, N° 1 (1979), p. 39-55.

Grisay, A. Rendement en français, notes et échecs à l'école primaire : les mirages de l'évaluation scolaire. Les Sciences de l'Education, N° 3 (1984), p. 107-148.

Grisay, A. Que peut-on "prescrire" en matière d'évaluation-bilan ? In: J.-M. De Ketele (Ed.). L'évaluation : approche descriptive ou prescriptive ? Bruxelles: De Boeck, 1986, p. 161-177.

De Ketele, J.-M. Objectifs terminaux d'intégration et situations d'interrogation : caractéristiques et exemples. In: J.-M. De Ketele (Ed.). L'évaluation en vue du conseil de classe : Documents de travail pour la formation. Louvain-la-Neuve: Laboratoire de Pédagogie Expérimentale, 1984.

De Landsheere, G. Dictionnaire de l'évaluation et de la recherche en éducation. Paris: P.U.F., 1979.

Laugier, H. et Weinberg, D. Recherches sur la solidarité et l'indépendance des aptitudes intellectuelles d'après les notes des examens écrits du baccalauréat. Paris: Chantenay (Imprimeur), 1938.



Mager, R. Measuring instructional results. Belmont (Ca): Pitman Learning Inc., 1984 - Traduction: Comment mesurer les résultats de l'enseignement. Paris: Bordas, 1986.

Monteil, J.-M. Effets de la valorisation et de la dévalorisation sur les performances scolaires dans une situation inter-groupe : étude expérimentale. Clermont-Ferrand: Université de Clermont-Ferrand, Laboratoire de Psychologie sociale, 1987.

Noizet, G. et Caverni, J.-P. Psychologie de l'évaluation scolaire. Paris: P.U.F., 1978.

Perrenoud, Ph. La fabrication de l'excellence scolaire dans l'enseignement primaire : du curriculum aux pratiques d'évaluation. Genève: Droz, 1984.

Perret-Clermont, A.-N. La construction de l'intelligence dans l'interaction sociale. Berne: Peter Lang, 1979.

Pfister, Ch. La validité de la note scolaire. Berne: Herbert Lang, 1975.

Piéron, H. Examens et docimologie. Paris: Presses Universitaires de France, 1969.

Schubauer-Léoni, M.-L. et Perret-Clermont, A.-N. Interactions sociales dans l'apprentissage de connaissances mathématiques chez l'enfant. In: G. Mugny (Ed.). Psychologie sociale du développement cognitif. Berne: Peter Lang, 1985, p. 225-250.

Tourneur, Y., Longo, M., Boulet, C., Dessailly, P. et Stievenart, M. Diagnostic des causes du retard pédagogique imputables au fonctionnement du système éducatif et proposition d'un ensemble de mesures préventives de l'échec scolaire dans l'enseignement primaire. Charleroi: CUNIC, 1981.

Vergnaud, G. L'enfant, la mathématique et la réalité. Berne: Peter Lang, 1981.