

L'ÉVALUATION DANS LE DOMAINE DE L'ÉDUCATION : TENDANCES

G. de LANDSHEERE
Université de Liège

RESUME :

Parallèlement au bond en avant que la recherche en éducation a fait ces dernières années, la théorie de l'évaluation a remarquablement progressé.

On comprend de mieux en mieux combien les évaluations normatives sont de peu d'utilité pour le pilotage des apprentissages. Pour aider l'élève à progresser, il faut entretenir chez lui le désir d'apprendre et lui indiquer ou mieux, lui faire découvrir où il en est dans son parcours vers le savoir, le savoir-faire ou le savoir-être (ce que font les épreuves critérielles), et en quoi consistent les écueils auxquels il se heurte. L'évaluation formative répond à ce besoin.

Les épreuves critérielles et, avec elles, l'évaluation des compétences minimales sont appelées à jouer un rôle croissant dans le processus éducatif.

Par ailleurs, on perçoit aussi de mieux en mieux le caractère superficiel de bien des évaluations des acquis. Elles se fondent encore trop souvent sur des interrogations concernant un cas isolé, présenté sous une seule forme, dans une seule conditions, et à un seul moment. La théorie de la généralisabilité s'efforce de lever cette hypothèque.

A côté de la science de l'éducation, la psychologie cognitive, a elle aussi fait des progrès considérables au cours de ces dernières décennies. Ils ont des implications importantes pour l'évaluation. Une telle constatation n'implique cependant pas que toutes les méthodes de construction des examens ou des tests que nous connaissons sont inadéquates, mais bien qu'il faut les appliquer à des modèles cognitifs bien choisis.

Enfin, on observe un fort mouvement en faveur du testage des performances effectives. Dire ce que l'on ferait n'assure pas qu'on est capable de le faire dans la réalité... L'évaluation sur la base de portefeuilles réunissant des échantillons de productions personnelles suscite actuellement beaucoup d'intérêt. Tant les portefeuilles que les observations directes de performances posent aussi des problèmes de validité et de fidélité des observations.

A côté des améliorations qui viennent d'être évoquées, bien d'autres restent à apporter, à commencer par une meilleure diffusion des progrès théoriques de la docimologie dans la pratique scolaire. La situation change, certes, mais trop lentement par rapport aux transformations que l'essor vertigineux des technologies provoque dans notre société.

SUMMARY :

Generally speaking, educational research has considerably progressed in recent time and evaluation too. It is better and better understood that normative evaluation is not a great help of monitor learning. Need for

achievement is a critical factor and the learner should be continuously informed of his situation in the domains of knowledge, skills and attitudes. The sources of learning difficulties must be diagnosed. Criterion referenced testing and formative evaluation are the answer.

Minimum competency testing is also at the order of the day.

There is a better awareness of the superficiality of much achievement testing. It often happens that only one facet is superficially tested. The theory of generalisability clarifies the issue.

Performance testing is an other issue. Among other things, it must be realized that multiple choice items have still a place in testing, specially in abstract subjects like mathematics or logics, but they must be geared on higher cognitive processes. Performance portfolios open new evaluation perspectives ; however, they are not necessarily free from bias or lack of representativity. As for direct observations of performances, they are often time consuming and lack of validity and reliability is not exceptional.

In recent decades, major advances in cognitive psychology have been implemented. Some of the ways it should influence educational evaluation are indicated.

A better dissemination of these innovations in school practice should be stimulated. Progress is achieved, but it is too slow as compared with the breath taking changes in our way of life and in technologies.

Parallèlement au bond en avant que la recherche en éducation a fait ces dernières années, la théorie de l'évaluation a remarquablement progressé.

On a de mieux en mieux perçu combien les évaluations normatives, - celles qui informent sur la place qu'un élève occupe par rapport aux autres -, sont de peu d'utilité pour le pilotage des apprentissages. Pour aider l'élève à progresser, lui dire que d'autres font mieux ou moins bien que lui ne l'avance pas beaucoup. L'important est d'entretenir chez lui le désir d'apprendre et de lui indiquer ou mieux, de lui faire découvrir où il en est dans son parcours vers le savoir, le savoir-faire ou le savoir-être (ce que font les épreuves critérielles), et en quoi consistent les écueils auxquels il se heurte. **L'évaluation formative** répond à ce besoin.

Les épreuves critérielles et, avec elles, l'évaluation des **compétences minimales** sont appelées à jouer un rôle croissant dans le processus éducatif.

Par ailleurs, on perçoit aussi de mieux en mieux le caractère superficiel de bien des questions posées. En effet, même s'il s'agit d'un problème, elles ne concernent trop souvent qu'un cas isolé, présenté sous une seule forme, dans une seule condition, et à un seul moment. La **théorie de la généralisabilité** s'efforce de lever cette hypothèque.

Enfin, une évaluation dépourvue de **validité prédictive** est sans grand intérêt. Savoir que l'étudiant connaît les principales règles de la grammaire d'une langue étrangère et réussit les exercices de contrôle à leur propos n'est utile que dans la mesure où l'on établit que l'acquis peut effectivement servir à la prise d'informations ou à la communication.

A côté de la science de l'éducation, la **psychologie cognitive** a fait des progrès considérables au cours de ces dernières décennies. Ils ont des implications importantes pour l'évaluation.

Esquissant une nouvelle théorie des examens, R.J. Mislavy (1990) écrit :

"Les élèves n'accroissent pas leurs compétences en accumulant simplement des habilités et des faits nouveaux, mais en donnant une forme nouvelle à leurs structures de connaissances, en automatisant les procédures, en comprimant l'information pour réduire

la charge de la mémoire, et en développant des stratégies et des modèles qui leur disent quand et comment les faits sont appropriés. Les types d'observation et les configurations de données qui reflètent les façons dont l'étudiant pense, réalise et apprend ne sont pas couverts par les examens traditionnels."

Une telle constatation n'implique pas que toutes les méthodes de construction des examens ou des tests que nous connaissons sont inadéquates, mais bien qu'il faut les appliquer à des modèles cognitifs bien choisis.

À côté des améliorations qui viennent d'être évoquées, bien d'autres restent à apporter, à commencer par une meilleure diffusion des progrès théoriques de la docimologie dans la pratique scolaire. La productivité du système éducatif global et le trajet scolaire de beaucoup d'élèves, en particulier, souffrent considérablement d'erreurs fondamentales que l'on continue à commettre dans le domaine de l'évaluation. Cette insuffisance est due à plusieurs facteurs dont les effets se conjuguent : carences dans la formation initiale des enseignants, faiblesse de leur formation continuée, manque d'information des décideurs, insuffisance notoire des investissements dans les activités de développement d'instruments d'évaluation, manque d'évaluation des rendements scolaires, esquive de l'obligation de rendre des comptes aux membres de la communauté éducative.

La situation change, certes, mais trop lentement par rapport aux transformations que l'essor vertigineux des technologies provoque dans notre société.

La forme de civilisation dans laquelle nous entrons ne sera apogée historique que dans la mesure où le droit de tous à profiter pleinement de ses apports sera respecté (ce qui exige de nouvelles formes de solidarité et où tous seront bien préparés à faire usage de ce droit) ce qui exige une éducation réussie. Une évaluation constructive est l'un des principaux moyens d'y parvenir.

Selon J. Stewart et coll. (1990, p. 19), les principes généraux guidant l'évaluation doivent être les suivants :

1. L'intérêt des élèves doit primer. L'évaluation doit être conçue et exécutée de façon à maximiser les bénéfices pour les élèves et à en minimiser les effets négatifs.
2. La première raison d'être de l'évaluation doit être de fournir une information permettant d'identifier les points forts et de guider les améliorations. En d'autres termes, l'évaluation doit suggérer des actions susceptibles d'améliorer le développement éducationnel des élèves et la qualité des programmes éducatifs.
3. Les résultats de l'évaluation ne doivent pas être utilisés pour porter des jugements ou prendre des décisions politiques susceptibles de nuire aux élèves ou à l'efficacité des enseignants.
4. Tout doit être mis en oeuvre pour assurer à tous une évaluation correcte.
5. La participation de la communauté éducative est essentielle pour la crédibilité et l'effet des procédures d'évaluation. Toutes les parties intéressées devraient avoir l'occasion de se faire entendre. L'auto-évaluation doit être le point de départ.

6. Il importe d'envisager soigneusement les effets que les pratiques d'évaluation peuvent exercer sur la motivation. Si ces pratiques ont un effet négatif (pour les élèves, les professeurs, les écoles ou le système éducatif), elles sont presque certainement indésirables.
7. Une évaluation correcte exige que l'on tienne compte de beaucoup de facteurs.
8. Dans l'évaluation des apprentissages cognitifs, il importe d'accorder beaucoup d'attention aux habilités complexes, telles que la compréhension des principes, l'application des connaissances et des habiletés à des tâches nouvelles (transfert), à l'examen, l'analyse, la discussion de questions et de problèmes complexes.
9. Il faut accorder plus d'importance au développement de l'individu et aux progrès accomplis, qu'à la façon dont il se situe par rapport aux autres.
10. Le choix des informations relatives à l'évaluation qui seront diffusées dans le public détermine le bénéfice ou la nuisance qui peut en résulter. C'est pourquoi les raisons, la sélection, la présentation et la diffusion des informations doivent respecter les principes ci-dessus.

I-L'ÉVALUATION DES ÉLÈVES

Psychologie cognitive et perspectives nouvelles

Tant pour le pronostic que pour les épreuves de rendement ou de diagnostic, les progrès récents de la psychologie cognitive ouvrent des perspectives nouvelles. Même si les notions théoriques qui vont être évoquées, de façon très simplifiée, ne peuvent pas encore déboucher aisément sur des pratiques quotidiennes, elles aident au moins à prendre conscience de problématiques fondamentales.

Au lieu de se contenter de constater qu'un élève, qui n'a pas trouvé la solution d'un problème, n'a pas été capable d'accomplir telle tâche, la psychologie cognitive veut savoir quelle démarche mentale explique le phénomène. Grâce surtout aux travaux de Piaget, tout éducateur sait que les possibilités d'un élève sont liées au stade de développement cognitif, affectif, psychomoteur qu'il a atteint.

Par une comparaison avec l'ordinateur, le traitement de l'information par le cerveau peut être décrit de la façon suivante :

- Les structures du système cognitif peuvent être activées par des stimuli reçus par les organes des sens ou par d'autres structures de connaissances (pensée) ; des informations sont ainsi *entrées*.
- *La mémoire à court terme* retient un nombre limité d'informations pendant une courte période de temps. Les faits isolés sont oubliés les premiers.
- *La mémoire à long terme* ou mémoire active stocke de façon codée et permanente les concepts et les constructions mentales dérivées de l'expérience

passée de la personne (événements, relations, processus, affects...) et dirige les opérations de tout le système de traitement de l'information. Cette base de données est le coeur du système cognitif. La capacité de cette mémoire est illimitée.

- *Le système musculaire* exécute les actes moteurs : lire, parler, courir...

L'extraordinaire performance du cerveau ne peut s'expliquer que par l'existence :

- de **procédures** tellement rapides qu'elles ne demandent plus beaucoup d'attention (automaticité) ;
- de **stratégies** permettant de sélectionner, planifier, piloter et, au besoin, modifier les actions (habiletés métacognitives) ;
- de **structures mentales (schémas)** qui mettent les faits et les habiletés en relation. Les schémas sont des structures abstraites qui représentent des ensembles significatifs d'informations stockées en mémoire (par exemple, tout ce qui constitue le déroulement normal d'un match de football).

L'apprentissage consiste en automatisation de procédures, en acquisition et en développement d'habiletés métacognitives, et en construction, révision et remplacement des schémas.

Par rapport aux "novices", les "experts" se caractérisent par le nombre croissant d'interconnexions qu'ils établissent entre un nombre de faits de plus en plus grand, par l'ampleur et la qualité des schémas qu'ils organisent et par une automatisation poussée des procédures, ce qui leur permet de s'attaquer à de nouveaux aspects des problèmes, de se placer dans des perspectives différentes.

Plus techniquement, on parle :

- de *connaissances déclaratives*, connaissances relatives aux propriétés des objets et des événements du monde : faits, concepts, formules...
- de *connaissances procédurales* : elles articulent les connaissances déclaratives en procédures et représentent les savoir-faire pratiques. Par le travail sur les connaissances déclaratives (processus métacognitifs) et aux feedbacks qui l'accompagnent, des règles se dégagent et des procédures se combinent de mieux en mieux en séquences. Dans les cas favorables, les procédures deviennent automatiques ; la place est ainsi laissée libre pour l'acquisition de comportements nouveaux, plus élaborés. Par exemple, celui qui, à la simple lecture d'une partition musicale, est capable de l'exécuter automatiquement peut alors consacrer l'essentiel de ses efforts et de son attention à la qualité de l'interprétation.
- de *schémas* : ce sont des structures cognitives qui spécifient les propriétés générales d'un objet ou d'un événement et abandonnent tout aspect contingent. Ils permettent à des objets d'être rattachés à des catégories générales (fonction d'abstraction) et exercent aussi des fonctions d'anticipation : ils guident la pensée dans des tâches de résolution de problèmes (exemple : prévoir le contenu d'un texte que l'on commence à lire). Les schémas s'organisent et se réorganisent par l'expérience et l'acquisition d'informations nouvelles.

- de *représentations* : ce sont des "théories personnelles", des systèmes de croyances relatifs à des phénomènes. C'est la façon dont un individu ou un groupe les comprend, se les explique. Par exemple, croire que l'électricité coule dans les fils comme l'eau dans les tuyaux, puisque, dans les deux cas, on parle de courant. Souvent tenaces, les représentations ne sont pas faciles à débusquer, ni à évaluer. Plus elles sont incorrectes et inadéquatement articulées, plus elles rendent difficiles de nouveaux apprentissages.

Implications pour l'amélioration des tests de performances scolaires

A la lumière des apports de la psychologie cognitive, H. Walberg (1991) estime que l'évaluation scolaire devrait remplir les fonctions suivantes :

- Mesurer la quantité de connaissances déclaratives que possède l'élève et identifier le type d'organisation qu'il utilise pour les stocker.
- Mesurer la vitesse d'exécution des tâches cognitives. Le taux d'attention requis peut être manipulé par l'introduction de tâches concurrentes ou par une augmentation de la difficulté des tâches.
- Apporter rapidement un feedback aux élèves et attirer leur attention sur les erreurs de processus ou de contenu.
- Diagnostiquer la source et le type des erreurs.
- Déterminer, par l'analyse des erreurs, le degré d'adéquation, d'étendue et de flexibilité des schémas.
- Identifier les stratégies de résolution de problèmes ; elles sont, en général, spécifiques à un domaine.
- Observer dans quelle mesure l'élève est capable de varier ses stratégies de résolution de problèmes lorsqu'il ne trouve pas directement les solutions.
- Identifier les stratégies métacognitives utilisées par l'élève; repérer celles qui l'aident et celles qui ne l'aident pas. Il importe de distinguer les habiletés métacognitives générales et les habiletés métacognitives spécifiques à un contenu.
- Pour évaluer la qualité des stratégies de résolution de problèmes, prendre pour référence la façon dont les experts perçoivent le problème, élaborent et choisissent un plan de résolution et mettent la stratégie arrêtée en oeuvre.
- Evaluer la qualité des réseaux sémantiques.
- Evaluer le degré d'automatisme des habiletés intellectuelles.

Conclusion

- Des techniques d'évaluation permettent de connaître les processus cognitifs de l'élève.

- L'utilisation des tests de connaissances à facettes s'impose (généralisabilité). En particulier, la façon de questionner doit être variée, de même que le niveau de difficulté. On peut ainsi estimer dans quelle mesure la façon de poser les questions influence les performances observées.

- L'évaluation doit être dynamique. On ne doit pas se contenter d'observer ce que les élèves sont capables de faire par eux-mêmes, à un moment donné. Il faut aussi déterminer jusqu'où ils peuvent aller lorsqu'on les met sur le chemin de la solution (étude du *développement proximal*, selon l'expression de Vygotsky, 1978).

Il s'impose, toutefois, de rester prudent à propos de toutes ces propositions. On ne dispose que de peu d'études statistiques à leur sujet. La validité de contenu, la généralisabilité des épreuves doivent être établies. De surcroît, les démarches proposées sont coûteuses.

EPREUVES DE PERFORMANCE OU EPREUVES DE RECITATION ? UN DEBAT FONDAMENTAL

La plupart des examens et des tests continuent à demander, sous forme plus ou moins déguisée, une simple récitation écrite (épreuves dites "papier-crayon") ou orales.

On classe aussi dans la catégorie *papier-crayon* les examens ou tests non verbaux dont les consignes sont données oralement et où les questions sont posées sous forme de dessins, de symboles, de photos parmi lesquels il faut faire un choix ou sur lesquels il faut indiquer une caractéristique.

Dans une première acception, on appelle *épreuves de performance* celles qui appellent des réponses motrices, manuelles ou non : manipulation d'objets, construction d'après un modèle, assemblage... Cette limitation au domaine moteur est cependant erronée, du moins dans la perspective actuelle. En effet, on considère plus généralement comme test de performance toute épreuve qui oblige l'élève à construire effectivement sa réponse, au lieu de n'avoir qu'à la choisir toute faite dans sa mémoire ou en appliquant une démarche de résolution de problème routinière et stéréotypée.

Ne relèvent donc pas des épreuves de performance des questions telles que :

- Quelle est la formule de l'acide sulfurique ? (Pure mémoire).
- Sachant que l'aire d'un triangle se calcule par la formule $\text{base} \times \text{hauteur} / 2$, quelle est l'aire d'un triangle de 25 cm de base et de 30 cm de hauteur ? (Compréhension).
- Calculez l'aire d'un triangle de 25 cm de base et de 30 cm de hauteur. (Application).

Dans la taxonomie de B.S. Bloom (domaine cognitif) et selon le sens conventionnel qui y est attribué aux trois termes *mémoire*, *compréhension*, *application*, ces trois questions font appel à des processus cognitifs inférieurs.

En revanche, bien que papier-crayon, la question à choix multiple suivante exige la mise en oeuvre de processus cognitifs supérieurs et relève donc bien des épreuves de performance :

Le parquet d'une chambre doit être fait de lattes entières. A cet effet, M. lattes de a cm sur b cm sont nécessaires. Combien faudrait-il de lattes mesurant x sur y cm ?

A. $\frac{Mab}{xy}$ B. $\frac{ab}{Mxy}$ C. $\frac{(a+b)M}{x+y}$ D. $\frac{ab.xy}{M}$ E. $\frac{Mxy}{ab}$

Ne relèvent pas non plus des épreuves de performance des examens où l'on peut répondre, pour l'avoir appris de mémoire, ce qu'il faudrait faire dans un cas donné, mais où l'on ne le fait pas effectivement. Exposer oralement ou par écrit la pédagogie progressiste que l'on pratiquerait si l'on était professeur ne prouve en rien qu'on est capable de la pratiquer effectivement, voire qu'il est possible de le faire étant données les limites circonstanciées qui existent dans la plupart des situations scolaires.

Depuis peu s'est développé un courant très fort en faveur d'examens qui portent sur la performance réelle, non seulement dans des matières relativement simples comme la dactylographie ou la capacité de régler un moteur, mais aussi à propos d'apprentissages plus abstraits et plus complexes. La difficulté réside dans la conception d'épreuves qui permettent d'évaluer de façon valide ce que le sujet est effectivement capable de faire (Linn, Baker et Dubar, 1991).

Trois grandes voies s'offrent en matière de testing de performance :

1. Pour les disciplines théoriques (philosophie, logique, algèbre, physique...), poser des questions qui exigent la mise en oeuvre de processus cognitifs supérieurs.
2. Faire effectuer la performance-cible (par exemple, faire une leçon de langue étrangère) devant les examinateurs.
3. Examiner des productions réelles à travers lesquelles la qualité de la performance peut être estimée. Il s'agit alors d'un *examen sur dossier* (le *portfolio* des Anglo-Saxons). Par exemple, pour faire valoir ses capacités pédagogiques, un professeur peut constituer un dossier qui réunit des préparations de leçons, du matériel didactique qu'il a confectionné, des enregistrements vidéo de ses leçons, des travaux d'élèves, des enregistrements de conversations avec des parents d'élèves, des articles publiés, etc.

Parmi les aspects qui menacent la validité des épreuves de performance, relevons (Stiggins, 1987) :

- Si le type de performance demandé est connu, le bachotage peut réapparaître.
- Un dossier ne se constitue pas naïvement. Si l'intéressé le prépare seul, il y placera sélectivement des témoignages à son avantage. Et si le dossier est constitué de façon négociée entre évalué et évaluateur, ce sont les meilleures pièces qui seront choisies de préférence. La solution théorique consiste à choisir

les témoignages au hasard en s'efforçant de constituer un échantillon représentatif des performances. C'est plus facile à dire qu'à faire !

- L'évaluation des performances peut, elle aussi, être biaisée par la perception qu'en ont les évaluateurs.
- Il est important d'établir si les performances observées sont réellement représentatives de la capacité ou de l'habileté.
- Les épreuves de performance sont censées mieux se prêter que d'autres à l'évaluation des processus cognitifs supérieurs : résolution de problèmes, esprit critique, raisonnement, métacognition... Si la tâche proposée est très familière au sujet, il peut l'exécuter de façon routinière, sans mettre en oeuvre les processus supérieurs.
- Les tâches doivent être significatives pour les sujets.
- Enfin, les épreuves de performance prennent beaucoup de temps, ce qui empêche souvent de proposer une quantité et une diversité de tâches suffisantes.

La théorie de la généralisabilité

Sans être inexacte, la façon dont on traite habituellement de la fidélité n'est en réalité, qu'une première approche, relativement grossière de la problématique.

Les résultats observés à un examen varient notamment :

- selon les questions posées (facette questions) ;
- selon le moment où l'épreuve est subie (facette occasions) ;
- selon les points de vue et la qualité des examinateurs (facette juges).

Par généralisabilité d'une mesure, on entend le degré auquel on peut, à partir d'une observation ou épreuve particulière, tirer des conclusions sur la valeur théorique recherchée.

Par exemple, à propos de la facette questions, il importe de considérer :

1. Le langage utilisé

- Influence du vocabulaire utilisé et de la syntaxe (lisibilité psychologique).
- Mode de présentation matérielle et psychologique. Par exemple, bien que portant exactement sur le même problème, les questions suivantes peuvent donner des résultats différents :

$$3 + 5 = ?$$

J'achète un crayon à cinq francs et un bonbon à trois francs, combien dois-je payer ?

2. La nature de la question

J. Cardinet aime à donner l'exemple suivant à propos du calcul de l'aire du triangle. Dans un groupe qui a étudié cette question pendant l'année scolaire, on obtient des résultats différents selon que l'on demande :

- Quelle est la formule de calcul de l'aire du triangle ?
- Quelle est la surface d'un triangle de 8 m de base et de 5 m de hauteur ?
- Voici un triangle et une règle graduée. Calculez la surface de ce triangle ?
- Même question que la précédente, mais la forme du triangle est telle que la perpendiculaire abaissée à partir du sommet tombe en dehors de la base.

Bref, connaître une formule ne suffit pas à témoigner d'une capacité réelle. Seules des performances concrètes, suffisamment variées dans leurs conditions pour constituer un échantillon représentatif des cas qui peuvent se présenter, permet une évaluation fidèle et valide.

A propos de cette facette "occasions", on peut notamment poser les questions suivantes :

- L'élève a-t-il eu effectivement l'occasion d'apprendre ce dont il va être question ?
- Cet apprentissage est-il récent ou déjà ancien ? Dans ce second cas, la connaissance a-t-elle été réactivée ou est-elle restée dormante ?
- Quel pouvait être le degré de fatigue de l'élève quand il a subi cette épreuve ?
- Une même question est-elle posée sous des formes diverses ?

Quant à la facette "juges", elle a maintes fois été évoquée par les docimologues : inconsistance des appréciations, effet de halo, effet de contraste, stéréotypie, effet de fatigue du correcteur ou de l'interrogateur...

L'évaluation de la compétence minimale

Toute épreuve finit par un constat de réussite ou d'échec. La compétence minimale marque la limite entre les deux. La définir objectivement n'est pas une mince affaire. Le mouvement actuel à propos des compétences minimales prend cependant une signification particulière. Malgré une prolongation importante de la scolarité dans les pays industrialisés, on constate que maints élèves quittent l'école sans maîtriser des habiletés de base comme la lecture, l'expression écrite et l'arithmétique élémentaire. C'est pourquoi apparaissent dans de nombreux pays des dispositions réglementaires subordonnant, par exemple, l'attribution d'un diplôme d'enseignement secondaire à la réussite d'une épreuve préliminaire établissant que l'élève maîtrise de façon suffisante les habiletés jugées essentielles.

FIXATION DE LA NOTE DE REUSSITE

Un examen ou un ensemble d'examens sont réussis ou non. La note de succès correspond au minimum de performance jugée acceptable, autrement dit à la compétence estimée minimale. Dans une perspective de la pédagogie de la maîtrise, on rencontre un

problème similaire quand on se demande à partir de quel moment l'élève a atteint cet état.

Plusieurs appellations désignent ce point critique : *note de réussite, note de coupure, seuil de réussite, niveau d'exigence minimale...* La fixation de cette note soulève presque toujours un problème redoutable. Comment, par exemple, déterminer à partir de quel moment précis quelqu'un connaît une langue étrangère ?

La question n'est simple que si l'on juge - là où c'est possible - en termes de tout ou rien. Ainsi, on peut décréter qu'un élève n'aura pas réussi aussi longtemps qu'il commet une ou plusieurs erreurs dans la multiplication des dix premiers nombres. On n'est bon pilote d'avion qu'à partir du moment où l'on réussit *tous* les atterrissages.

Dès qu'une part d'erreur est jugée admissible et si l'acquisition à évaluer est complexe, la note de coupure devrait être nuancée sous forme de fourchette à l'intérieur de laquelle on réussit. On parle alors de *zone de coupure*.

La note de réussite est tantôt *absolue* et déterminée a priori (on fixe la note avant l'épreuve, par rapport à un objectif à atteindre) ou a posteriori (on prend pour référence les résultats d'un groupe extérieur ou ceux du groupe même).

1. LES DECISIONS EMPIRIQUES

1.1 NOTE DE REUSSITE AUX EXAMENS TRADITIONNELS

L'une des règles les plus fréquentes dans notre pays était (et est parfois encore) d'obtenir au moins 50 % dans chaque branche et 60 % pour l'ensemble. Aucune justification objective de ces seuils n'est fournie, pas plus d'ailleurs que les raisons de placer ou non toutes les branches sur le même pied.

Parmi les démarches plus systématiques pour fixer la note de coupure, Ebel (2) propose les suivantes :

a) Pour les tests à choix multiple

Partant du principe que, dans un test à choix multiple bien construit, un élève ignorant peut obtenir un score correspondant à des réponses au pur hasard (50% pour des réponses juste-faux ; 25% pour des réponses à quatre choix, etc.) et qu'un élève très fort devrait arriver à peu près au maximum, Ebel estime que la note de coupure peut, dans ce cas, être calculée de la façon suivante :

Choix	Quatre choix	Juste - Faux
Nombre d'items	100	100
Score attendu si on répond au hasard	25	50
Score moyen idéal	$(100 + 50)/2 = 75$	$(100 + 25)/2 = 62,50$
Score de coupure	$(50 + 75)/2 = 62,50$	$(25 + 62,50)/2 = 43,75$

Le score moyen idéal se situe donc à mi-chemin entre le score maximum possible et le score attendu si l'on répond au pur hasard.

Le score de coupure se situe à mi-chemin entre le score moyen idéal et le score attendu si l'on répond au pur hasard.

b) Pour les examens traditionnels

Le score de réussite est fixé arbitrairement à un pourcentage donné ; ce score est ensuite ajusté en fonction des performances observées.

Exemple : on estime que 3/4 des élèves devraient fournir des réponses correctes, à condition qu'au moins 60% et qu'au plus 80% d'entre eux dépassent ce score.

Si moins de 60% des élèves arrivent à 75% de réussite, le score de réussite est fixé exactement à la moitié de la distance entre le score correspondant à 75% et le score qu'au moins 60% des élèves dépassent.

Si plus de 80% des élèves dépassent le score de réussite, ce score est alors situé à mi-chemin entre le score correspondant à 75% et le score que 80% des élèves dépassent.

Ebel précise (p. 341) : "Les chiffres de 60 et 80% des sujets sont pris purement à titre d'exemple. Selon les circonstances, on peut souhaiter les augmenter ou les diminuer. Le but poursuivi en fixant les valeurs est de maintenir le niveau de connaissances et le pourcentage de réussite dans ce qui semble des limites raisonnables, tout en tenant aussi raisonnablement compte de l'erreur sur la mesure. Cette méthode n'est pas souvent utilisée et, pourtant, elle offre le mérite d'apporter une solution rationnelle à un problème difficile."

1.2 NOTE DE REUSSITE AUX EPREUVES DE COMPETENCE MINIMALE

Dans la pédagogie de la maîtrise, le seuil de réussite exigé se situe, en principe, à 100%. Cette exigence est souvent modérée, ne fût-ce que pour tenir compte des erreurs fortuites, accidentelles. En général, on accepte 90, voire 80%, mais pas moins.

En évaluation normative, il arrive que le seuil fixé corresponde à la performance moyenne d'un groupe de référence. Par exemple, V. de Landsheere (1988, p. 124) rapporte que, dans plusieurs Etats américains, l'attribution du diplôme de fin d'études secondaires est subordonné à l'obtention, à un test en capacité de lecture, d'une note au moins égale au niveau moyen des classes de 3^e année.

V. de Landsheere écrit très justement (1988, p. 130) : "Fondamentalement, un seuil d'exigence unique est impossible à fixer parce que le degré de perfection voulu varie en pratique selon l'objectif poursuivi et la notion de l'apprentissage à réaliser. Rares sont les cas simples comme : on sait rouler à bicyclette lorsque, sauf accident, on n'en tombe jamais et lorsqu'on est capable de démarrer et de s'arrêter à volonté. Dans ce cas, le niveau d'exigence est pratiquement de 100%. Comment appliquer une règle similaire à <Savoir jouer du piano> ? Comment justifier que, dans les bulletins scolaires, on demandait une note d'au moins 50% aussi bien en langue maternelle, en

mathématiques... qu'en gymnastique et en dessin ? Que signifie d'ailleurs savoir à moitié dessiner ?"

Bref, aucune des démarches qui viennent d'être décrites n'est pleinement satisfaisante. Tout au plus apportent-elles un peu moins d'anarchie dans les décisions.

II. VERS PLUS D'OBJECTIVITE

Depuis les années 70, de nombreuses recherches ont porté sur les méthodes scientifiques de détermination de la note de coupure et sur le taux d'erreurs que l'on commet quand on les applique. En 1986, Beck relève plus de 35 de ces méthodes. On trouve dans l'ouvrage que V. de Landsheere (1988) a consacré à la compétence minimale, la présentation et une analyse critique des quinze méthodes qui ont le plus retenu l'attention. Phénomène qui doit beaucoup faire réfléchir les éducateurs : selon la méthode choisie, le taux d'échec varie de 9 à 78% !...

Nous ne décrivons que la méthode d'Angoff (1971), parce qu'elle réunit un ensemble de qualités largement reconnues et qu'elle semble la plus sage.

1. On demande à un groupe d'éducateurs expérimentés d'indiquer, pour chaque question d'un examen ou d'un test, quel pourcentage de chances de répondre correctement a un élève minimalement compétent, dans le groupe à examiner.

10 - 20 - 30 - 40 - 50 - 60 - 70 - 80 - 90 - 100

2. Chacun des juges note ses appréciations sur le tableau suivant :

Juge n° 1

Pour 100 élèves, à la limite de la compétence, combien, à votre avis, répondraient correctement aux questions posées ?

Questions	Pourcentage	Pourcentage / 100
1	80 %	0,8
2	40 %	0,4
3	50%	0,5
4	0 %	0
5	20 %	0,2
Note minimum de réussite : 1,9		

Pour obtenir la note de coupure sur l'ensemble des juges, on calcule la moyenne.

Dans cette démarche, la suggestivité ne cherche pas à se cacher, ce qui incite en permanence à la pondération dans les décisions.

Saunders (1984) propose un perfectionnement intéressant de cette méthode.

Après avoir fait les calculs selon Angoff, les juges se voient offrir trois occasions de réviser leurs appréciations :

1. Après avoir pris connaissance des notes de coupure auxquelles les autres juges ont abouti.
2. Après avoir, en plus, pris connaissance de la distribution des scores obtenus par les élèves qui ont subi le test.
3. Après avoir étudié les statistiques descriptives (moyenne, médiane, marge de variation, écart type), élaborées à partir des notes de coupure fixées au stade précédent (principalement pourcentage d'élèves qui réussiraient).

La note de réussite est finalement fixée sur la base des positions prises au stade 3, lors d'une discussion entre tous les juges.

III. CONCLUSION

Même si elles présentent un indéniable progrès par rapport à l'exigence traditionnelle d'au moins 50 ou 60% dans chaque branche, les démarches qui viennent d'être décrites restent tâtonnantes.

Une préoccupation majeure doit toujours rester à l'esprit : la fonction première de l'évaluation n'est pas de classer les élèves, mais bien d'informer sur ce qu'ils ont appris effectivement et sur ce qui leur reste à faire pour être au moins minimalement compétents.

Cette notion de "compétence minimale" est, à la limite, aussi insaisissable que celle de la "note vraie", notamment parce qu'elle varie en fonction des contextes. Être minimalement compétent pour piloter un petit avion de tourisme ne qualifie pas pour prendre les commandes d'un gros avion de ligne... De surcroît, même s'il n'est pas sans utilité, un constat d'incapacité n'offre pas d'intérêt réel pour l'éducateur s'il ne s'accompagne pas d'un diagnostic clair ouvrant la voie à la remédiation.

LE TESTING ADAPTATIF

Les tests dits adaptatifs permettent d'individualiser l'examen.

Un tel test est constitué de questions ordonnées par ordre de difficulté croissante. Soit au hasard, soit en fonction de sa compétence estimée (éventuellement par des examens précédents), l'élève répond à une première question ou à un premier ensemble de questions. Le choix des questions qui suivront est commandé par la qualité des réponses initialement données, et ainsi de suite. L'épreuve s'arrête quand la stabilité des réponses à un niveau donné permet de penser qu'il est bien celui que l'élève a atteint effectivement.

Grâce à l'ordinateur, les tests adaptatifs semblent appelés à jouer un rôle important dans un avenir proche.

La construction artisanale d'une telle épreuve se fait par tâtonnements et jugements. En s'appuyant sur son expérience et sur des résultats antérieurement observés, l'éducateur détermine empiriquement une hiérarchie de difficulté (qui réside à la fois dans le contenu objectif de la question et dans sa forme). Il existe aussi des techniques statistiques sophistiquées pour construire ce type d'instrument. Les plus connues procèdent de deux modèles d'interactions entre la compétence du sujet et la difficulté des questions : le modèle d'analyse hiérarchique de Guttman (in MATALON, 1975) et le modèle de Rasch (in LECLERCQ, 1980).

B. L'EVALUATION DU SYSTEME SCOLAIRE

I. LE PILOTAGE DU SYSTEME

Observer de façon systématique et, idéalement, continue, comment le système éducatif se comporte présente un intérêt incontestable. Une telle observation est appelée *pilotage* ou *monitoring*.

En technologie, un moniteur est un dispositif de surveillance qui déclenche une alarme en cas de trouble de fonctionnement. En matière d'enseignement, on distingue trois types de pilotage :

- Le *pilotage administratif* qui veille au respect des dispositions réglementaires : utilisation correcte des crédits alloués, engagement d'un personnel possédant les titres requis, contrôle du nombre d'élèves inscrits... Cette surveillance existe depuis longtemps.
- Le *pilotage formatif*, axé sur les individus. Par exemple, l'administration périodique de tests critériels couvrant l'essentiel d'un programme d'études permet d'identifier les problèmes qui peuvent se poser dans un environnement scolaire donné et d'en chercher la solution.
- Le *pilotage du rendement scolaire*. Il est généralement réalisé à l'aide de tests normatifs informant de façon comparative sur ce que "produisent" les écoles et les classes. Une évaluation centrée sur les objectifs est aussi possible. Les considérations qui suivent concernent essentiellement ce troisième type.

On peut contrôler la productivité en savoirs, savoir-faire et savoir-être d'un système éducatif, soit indépendamment des objectifs du curriculum (on dresse le bilan de tous les acquis identifiables), soit en liaison avec les objectifs que l'enseignement s'est explicitement donné pour mission d'atteindre.

C'est dans cette seconde hypothèse que le pilotage s'indique d'abord. On veut savoir si l'on avance bien dans la direction choisie et, à cette fin, on prélève des indicateurs qui deviendront éventuellement des clignotants si des problèmes particuliers se posent. Si

l'évaluation est répétée périodiquement, il devient possible de distinguer des évolutions et des tendances.

L'importance politique du pilotage est indiscutable. Si, en outre, les résultats des observations sont publiés et largement diffusés, on fait un pas important dans la gestion démocratique.

Dans une note inédite de 1990, G. de Landsheere et M. Crahay (Université de Liège, Service de pédagogie expérimentale) précisent que le pilotage devrait établir dans quelle mesure le système éducatif répond à trois exigences : la cohérence, la qualité et l'équité.

La cohérence existe :

- si les objectifs poursuivis répondent aux besoins et aux options privilégiées du corps social ;
- si le curriculum implanté - c'est-à-dire, ce qui est effectivement enseigné en classe - correspond aux objectifs fixés ;
- si la nature et les modalités des évaluations sont fonction de ces mêmes objectifs.

Le critère d'équité exige notamment que les ressources disponibles profitent également à tous les élèves. Il convient, en particulier, d'établir si la qualité et le niveau des acquis ne sont pas liés davantage à l'établissement fréquenté qu'aux capacités réelles de ceux qui le fréquentent. En démocratie, le système éducatif a pour idéal l'égalité de réussite, étant entendu que les modalités de celle-ci peuvent varier selon les aptitudes, les besoins et les intérêts de chacun.

Bref, bien des questions appellent des réponses claires et fondées. Qu'apprennent réellement les élèves ? Le système éducatif est-il équitable ? Les programmes répondent-ils aux exigences du monde d'aujourd'hui et de demain ? Les méthodes d'enseignement sont-elles adéquates ? Ces questions sont évoquées dans le tableau suivant.

Récapitulation des questions :

	Cohérence	Qualité absolue	Qualité relative	Equité
Objectifs des programmes	L'enseignement répond-il aux besoins des élèves ?	Les programmes sont-ils assez mais pas trop exigeants ?	Les élèves sont-ils aussi bien formés que ceux d'autres pays ?	Les programmes sont-ils conçus pour assurer l'égalité des chances ?
Pratiques réelles dans les classes	Correspondent-elles aux objectifs à poursuivre ?	Le curriculum implanté est-il valide ?	Quelles pratiques s'avèrent plus efficaces que d'autres ? Quelles leçons tirer de ce qui se passe en classe ?	L'enseignement n'est-il pas biaisé en fonction des origines sociales ou ethniques, du sexe, de l'implantation géographique ?
Acquis des élèves	Le profil de compétences obtenu est-il le profil souhaité ?	Les acquis sont-ils suffisants pour exercer la citoyenneté responsable ?	Comme se situe : - l'élève moyen, - l'élève du groupe supérieur, par rapport à ceux d'autres pays ?	Observe-t-on d'importantes différences entre les catégories ci-dessus ? Quels sont les taux d'échecs ?
Evolution	En perçoit-on une ?	Le niveau du rendement baisse-t-il ?	Le système évolue-t-il mieux ou moins bien que celui d'autres pays ?	Le système évolue-t-il vers plus de démocratie, moins d'inégalité ?

Conclusion

Les uns après les autres, les pays industrialisés mettent en place, parfois avec des moyens considérables, des systèmes de pilotage et en publient les principaux résultats dans les médias. Les responsables de l'enseignement rendent ainsi des comptes de plus en plus circonstanciés à la communauté éducative. Mais cela ne suffit pas. Il faut, en effet, que les évaluations, surtout si elles conduisent à des conclusions négatives, soient suivies d'efforts d'amélioration. Aussi, pour que le pilotage produise vraiment les effets souhaités, il importe que les enseignants soient associés à sa conception et à sa réalisation.

"Si les enseignants n'ont pas participé à l'identification des problèmes ou des inadéquations de l'école et à la formulation de solutions réalistes et prometteuses, ils ne croiront pas nécessairement que des problèmes existent ou qu'une solution proposée est susceptible d'améliorer les pratiques courantes. (...) Un changement significatif du fonctionnement de l'école exige un changement des attitudes et des pratiques des enseignants et aussi d'autres parties du système d'enseignement et d'apprentissage. Les maîtres enseignent ce qu'ils comprennent, ce qu'ils croient important et ce qu'ils pensent pouvoir transmettre avec succès. Il est clair que les enseignants sont les agents les plus déterminants d'une réforme scolaire." (Tyler, 1983, p. 462)

II. LES INDICATEURS

Mesure destinée à servir le pilotage, un indicateur est une statistique directe et valide informant sur l'état et les changements d'ampleur et de nature, au cours du temps, d'un phénomène sociétal jugé important. En éducation, cette mesure informe notamment sur la santé et la qualité du fonctionnement du système, sur ce que connaissent les élèves, ce qu'ils sont capables de faire, et sur l'évolution positive ou négative de ces conditions (d'où l'intérêt à accorder à l'observation de caractéristiques durables) et sur des différences significatives qui peuvent exister entre des aires géographiques ou des institutions à un moment donné.

Dans de bonnes conditions, les indicateurs sont le reflet de la politique éducative et mettent éventuellement sur la voie de modifications à y apporter. Ils doivent être robustes, c'est-à-dire ne pas être trop sensibles à des fluctuations accidentelles. Si leur signification est facilement comprise par tous les membres de la communauté éducative, le dialogue entre les politiques, les citoyens et les enseignants est facilité.

Toutes statistique relative à l'éducation n'a pas valeur d'indicateur. Ainsi, connaître le nombre d'écoles que compte une région n'apprend rien sur la qualité de ce qui s'y fait. Un tel chiffre n'acquiert une signification que si un point de référence permet de l'interpréter : évolution d'une même statistique, comparaison avec des mesures similaires opérées ailleurs...

Concevoir des indicateurs est donc une opération où interviennent des considérations politiques et techniques. On attend qu'ils soient importants, valides, faciles à prélever, crédibles, et mettent sur la voie des décisions correctives à prendre.

En particulier, la validité interne d'un indicateur est menacée :

- si l'on n'opère qu'une mesure partielle d'un phénomène multidimensionnel ;
- si le phénomène à évaluer est défini de façon inadéquate ;
- si la variable substituée à une autre pour des raisons de facilité n'est pas valable ;
- si l'objet de la mesure a changé au cours du temps ;
- s'il y a eu "corruption". Cette falsification se produit, par exemple, si l'enseignant souffle les réponses aux questions d'un test.

Exemples

La consultation de la littérature spécialisée et des rapports de pilotage de plus en plus nombreux montre, si besoin en était, que des centaines de caractéristiques des systèmes éducatifs peuvent être observées. Les indications suivantes n'ont donc que valeur d'exemples.

Indicateurs relatifs aux élèves

a) *Cognitifs*

Ils concernent le plus souvent les performances scolaires :

- Scores obtenus à des tests nationaux ou internationaux
- Epreuves de compétences fonctionnelles
- Capacité de raisonner et de résoudre des problèmes
- Capacité d'analyser l'information

Etc.

b) *Socio-affectifs*

- Attitude vis-à-vis du travail, de la société...
- Satisfaction éprouvée à l'école
- Qualité de la relation avec les enseignants
- Taux de pauvreté
- Nombre d'heures consacrées à la télévision
- Agressivité - violence
- Alcoolisme, drogue

Etc.

Indicateurs relatifs à l'école

- Climat de l'établissement
- Groupement des élèves (colonnes de classes, groupes de niveau...)

- Pourcentage d'élèves connaissant imparfaitement la langue d'enseignement
- Nombre d'heures d'enseignement par discipline et par année
- Structure d'encouragement à la relation avec le monde extérieur
- Association des parents à la vie de l'école
- Taux de redoublement
- Taux d'abandon des études
- Taux d'obtention du diplôme
- Durée moyenne pour faire un cycle d'études

Indicateurs relatifs au coûts

- Pourcentage du PNB consacré à l'enseignement
- Coût par élève
- Frais de fonctionnement
- Contribution financière des parents
- Etc.

Indicateurs relatifs aux enseignants

- Pyramide des âges
- Sexe
- Qualification (études faites)
- Nombre d'années d'expérience
- Ponctualité - absences - taux de départs volontaires
- Rémunération
- Etc.

Le projet international de l'OCDE

Le Centre pour l'innovation et la recherche en éducation (CERI) de l'Organisation pour la coopération et le développement économiques (OCDE) a lancé, en 1988, un projet coopératif de mise au point et de prélèvement périodique d'indicateurs de l'enseignement aux niveaux national et international. En particulier, des indicateurs internationaux devraient aider à définir et à adapter les politiques dans les pays membres.

Les groupes d'indicateurs retenus sont :

- Les taux de scolarisation, les trajectoire des élèves (cursus) et les sorties du système éducatif.
- Les acquis des élèves.
- Les établissements scolaires, leur environnement et les systèmes éducatifs.
- Les coûts et les ressources des systèmes d'enseignement
- Les attitudes et les attentes des utilisateurs du système.

Les problèmes

Des problèmes souvent difficiles se posent à propos des indicateurs, notamment :

- Le manque de comparabilité des mesures de rendement scolaire.
- Limiter les mesures aux connaissances et aux habilités cognitives laisse dans l'ombre d'autres effets possible de l'éducation, tant pour l'individu que pour la société.
- Certaines caractéristiques comme la qualité des enseignants sont difficiles à évaluer massivement ; des variables substitutives comme le type de formation ou l'ancienneté n'ont qu'une faible validité.
- Plus les indicateurs sont globaux, plus des phénomènes importants risquent d'être masqués.
- La publication d'indicateurs a une signification politique et il n'est pas rare que des pressions visant à orienter les choix ou la présentation des résultats se manifestent.
- Plus une politique ou un programme d'éducation sont imprécis, plus il est difficile de choisir des indicateurs représentatifs.
- Les informations doivent être publiées le plus rapidement possible, car à mesure que le temps passe, la signification d'un indicateur peut changer.
- Même conçus et collectés avec soin, les indicateurs n'apportent pas des informations d'une signification évidente. Celle-ci varie selon les valeurs auxquelles on adhère et aussi selon les options pédagogiques. Les attentes vis-à-vis du système éducatif diffèrent d'ailleurs selon les groupes : élèves, enseignants, familles, entreprises, autorités religieuses....

En conclusion

L'intérêt des indicateurs de l'enseignement est grand. Il reste à en faire un bon usage.

De même que les résultats d'un examen ne suffisent pas à donner une image fidèle d'un individu, de même les informations apportées par les indicateurs doivent être complétées, nuancées, encadrées par d'autres évaluations, générales ou particulières, et aussi par des comparaisons et par les avis et les opinions des membres de la communauté éducative.

BIBLIOGRAPHIE

- AUBRET, J., *Reconnaissance et évaluation des acquis : le portefeuille de compétence*. Paris, Presses Universitaires de Paris VIII, 1992
- BERK, R.A. A consumer's guide to setting performance standards on criterion-referenced tests, *Review of Educational Research*, 1986, 56, 1, 137-172

- DE LANDSHEERE, G., *Dictionnaire de l'évaluation et de la recherche en éducation*, Paris, PUF, 1992, 2e éd.
- DE LANDSHEERE, V., *Faire réussir - Faire échouer. La compétence minimale et son évaluation*, Paris, PUF, 1988
- DE LANDSHEERE, V., *L'éducation et la formation. Science et pratique*, Paris, PUF, 1992.
- EBEL, R.L. *Essentials of Educational Measurement*, Englewood Cliffs, Prentice Hall, 1979, 3e ed.
- GLASER, R., Instructional technology and the measurement of learning outcomes. Some questions, *American Psychologist*, 1963, 18, 519-521
- LECLERCQ, D., Computerized tailored testing, *European Journal of Education*, 1980, 5,3.
- LINN, R. L., BAKER, E.L. et DUBAR, S.B., Complex, performance-based assessment : expectations and validation criteria, *Evaluation Comment*, Hiver 1991-1992.
- MATALON, B., *L'analyse hiérarchique*, Paris, Gauthier-Villars, 1975.
- MISLEVY, R.J., Foundations of a new test theory. In N. FREDENKSEN et I. BEJAR (Eds.), *Test theory for a new generation of tests*, Lawrence Erlbaum, 1990.
- STEWARD, J. et al., *Tomorrow's standards. The report of the ministerial working party on assessment for better learning*, Wellington, Ministère de l'éducation, 1990.
- STIGGINS, R.J., Design and development of performance assessment, *Education Measurement : Issues and Practice*, 1987, 6, 3, 33-42.
- TYLER, R., *A place called school. Phi, Delta, Kappa*, 1983, 3, 462-464.
- VIGOTSKY, L.S., *Mind in society : the development of higher cognitive processes*, Cambridge, Mass., Havard University Press, 1978.
- WALBERG, H., *The implications of cognitive psychology for measuring school achievement*, Chicago, University of Illinois, 1991, ronéotypé, pp. 19-22.