

# Les enseignants peuvent-ils utiliser l'évaluation pour améliorer l'apprentissage ?

**Paul BLACK**

School of Education  
King's College London  
Cornwall House Annex  
Waterloo Road  
London SE1 8WA, United Kingdom.

*(Traduit par Maurice Chastrette)*

### Résumé

*De nombreux problèmes ont empêché le développement de l'évaluation formative, conçue comme moyen pour améliorer l'apprentissage. En premier lieu, la compétition entre les fonctions d'évaluation et de contrôle conduit souvent à une prédominance d'une évaluation sommative faite par contrôle externe ; cette prédominance peut ruiner les efforts faits par les professeurs pour améliorer l'apprentissage. D'autre part, il est difficile de mettre en œuvre des méthodes satisfaisantes et pratiques pour rassembler l'information provenant en «feedback» de tous les élèves d'une classe type. On peut également penser qu'une amélioration de ce feedback mettrait à jour un si grand nombre de besoins chez les élèves que de nouvelles méthodes d'enseignement seraient nécessaires pour y faire face. Développer l'autocontrôle constitue probablement une part essentielle de toute stratégie destinée à améliorer l'évaluation formative. Seule une réforme systémique peut permettre d'affronter tous ces problèmes, réforme selon laquelle les méthodes pouvant conduire aux différents objectifs de l'évaluation seraient conçues de façon qu'elles se renforcent mutuellement.*

**Mots clés :** *contrôle et apprentissage, évaluation formative, évaluation sommative, feedback des élèves, réforme systémique.*

## **Abstract**

*The development of formative assessment so that it can make its full contribution to improved learning has been inhibited by several problems. One problem is the tension between the different functions of assessment and testing, which often leads to dominance by summative assessment in the form of external testing. This dominance can damage the efforts of teachers to improve learning. Another problem is the difficulty of developing adequate and practicable methods of collecting feedback information from all pupils in a typical classroom. It is also argued that better feedback would expose such a wide range of pupil needs that new teaching methods would have to be implemented in order to meet them. It is suggested that developing self assessment by pupils may be an essential part of any strategy for improving formative assessment. These various problems can only be tackled by a systemic approach to reform in which methods for achieving the various purposes of assessment are designed in such a way that they support one another.*

**Key words** : testing and learning, formative assessment, summative assessment, pupils' feedback, systemic reform.

## **Resumen**

*El desarrollo de la evaluación formativa en vista de contribuir al mejoramiento del aprendizaje ha sido inhibido por diferentes problemas. Uno de ellos es la tensión entre las funciones de evaluación y de control, que conduce frecuentemente a un predominio de la evaluación sumativa bajo la forma de control externo. Este predominio puede constituirse en un obstáculo a los esfuerzos de los profesores para mejorar la enseñanza. Otro problema es la dificultad de desarrollar métodos adecuados para recolectar informaciones en «feedback» provenientes de todos los alumnos de una clase tipo. En este artículo se discute igualmente el hecho que mejorando el feedback, uno haría emerger una tal cantidad de necesidades de los alumnos que sería necesario implementar nuevos métodos de enseñanza para satisfacerlas. Se sugiere que una autoevaluación por parte de los alumnos puede tener un lugar esencial en toda estrategia destinada a mejorar la evaluación formativa. Estos diferentes problemas no pueden ser afrontados que por una reforma sistémica, en la cual los métodos para lograr los diferentes objetivos de la evaluación sean concebidos de manera que se refuercen mutuamente.*

**Palabras claves** : control y aprendizaje, evaluación formativa, evaluación sumativa, feedback de los alumnos, reforma sistémica.

## L'IMPORTANCE DE L'ÉVALUATION FORMATIVE

*«Promouvoir l'apprentissage est le principal but de l'école. L'évaluation se trouve au cœur de ce processus. En effet, elle peut fournir un cadre dans lequel les objectifs pédagogiques peuvent être fixés, et les progrès des élèves enregistrés et exprimés. Elle peut fournir une base pour la planification des étapes suivantes de l'enseignement, en réponse aux besoins des élèves. En facilitant le dialogue entre enseignants, elle peut améliorer leur qualification professionnelle et aider l'école à renforcer l'ensemble de l'apprentissage au long du curriculum.» (DES, 1987)*

Le message que cet article veut apporter est qu'une bonne évaluation formative peut constituer un outil puissant pour améliorer la qualité de l'apprentissage, alors qu'elle est en général sérieusement sous-développée à l'école. Il existe donc d'immenses possibilités d'amélioration ; cependant pour les saisir il ne suffit pas d'apporter de modestes compléments aux travaux existants.

Une caractéristique particulière de l'évaluation formative est que l'information liée à l'évaluation est utilisée à la fois par l'enseignant et par les élèves, dans le but de modifier leur travail et de le rendre plus efficace. Amasser de l'information est sans grand intérêt si l'on ne peut pas s'en servir dans l'action, et, puisque l'information liée à l'évaluation ne manquera pas de révéler l'hétérogénéité des besoins de la classe concernant l'apprentissage, l'action qui sera nécessaire devra comporter, sous une forme quelconque, un enseignement différencié. Ce point a été nettement souligné par Perrenoud (1991) dont la phrase : *«il y a en chacun de nous le désir de ne pas connaître les choses sur lesquelles on n'a aucune prise»*, implique en outre que la conscience de notre incapacité à agir en fonction d'une information peut être une raison de ne pas rassembler cette information.

Si l'on veut que l'éducation réalise pleinement les bénéfices potentiels de l'évaluation par les professeurs, alors il faut faire avec eux un important travail pour mettre en place une évaluation formative efficace dans les programmes d'enseignement. Un tel travail implique nécessairement des programmes à long terme pour modifier les perceptions et les rôles des enseignants aussi bien que des élèves, en insistant particulièrement sur le développement chez les élèves de la prise de conscience de la situation et de leur responsabilité dans l'auto-évaluation.

## ÉVALUATION FORMATIVE ET SOMMATIVE

Un des obstacles principaux au développement de l'évaluation formative est qu'elle ne remplit que l'une des trois fonctions de l'évaluation, c'est-à-dire :

- l'assistance directe à l'apprentissage,
- la certification de chaque élève individuellement,
- la responsabilité des institutions et des enseignants par rapport à la société (Black, 1993a ; Haney, 1991).

La deuxième et la troisième de ces fonctions influencent toutes les deux la première. La fonction de responsabilité sociale peut affecter le statut et l'emploi des enseignants, tandis que la fonction de certification influe sur les chances de leurs élèves dans la vie. Bien que la fonction de responsabilité sociale ne nécessite pas de données sur les individus, elle est souvent remplie dans la pratique en utilisant des données de certification, obtenues à partir de l'évaluation de l'ensemble des étudiants. La plus grande partie des travaux d'évaluation, que ce soit dans la pratique ou dans la recherche-développement, a été consacrée aux fonctions de certification et de responsabilité sociale, aux dépens de la fonction formative.

Le contrôle externe peut prédominer dans le travail en classe, et ainsi déformer l'enseignement à un point tel que les conditions d'une bonne évaluation formative ne sont pas remplies. Les enseignants croient que préparer les élèves à un test spécifique, en se concentrant étroitement sur ses exigences et sur ses méthodes, va augmenter leurs succès. Les effets nuisibles sur l'enseignement d'un contrôle externe étroit sont bien connus : l'apprentissage se calque sur le contrôle en se centrant sur les aspects qui sont facilement testés, et par suite le niveau du travail en classe baisse, les élèves doivent travailler à un rythme trop rapide pour un apprentissage efficace, les méthodes créatives et innovantes, de même que certains contenus thématiques, sont abandonnés (cf. Black, 1993a). Ces tendances démotivent de nombreux élèves, en récompensant ceux qui travaillent dans le cadre limité des démarches qui permettent de réussir les tests.

Dans ce contexte, il n'est pas surprenant que beaucoup de professeurs considèrent l'évaluation avec suspicion et aversion. Il est difficile de trouver des exemples d'utilisation de l'évaluation formative comme composante forte d'un programme d'apprentissage efficace. L'image générale qui émerge des études faites sur les pratiques est celle d'une évaluation négligée. Ceci a été bien décrit pour l'enseignement scientifique en Australie (Butler et al., 1980), en Écosse (Black H., 1986), en Angleterre (Hodson, 1986) et aux États-Unis (Rudman, 1987). Dans leur revue des pratiques d'évaluation des professeurs de sciences aux États-Unis, Raizen et al.

(1989) expliquent la pauvreté de l'évaluation en classe principalement par le fait que l'évaluation est négligée dans la formation des professeurs.

Une revue des pratiques à l'école primaire en France (Grisay, 1991) examine les critères de passage dans la classe supérieure. La conclusion est que les critères utilisés sont «*virtuellement invalides selon les standards externes*» et que les examens de fin d'année sont «*encombrés de questions sélectives tendant à mesurer la maîtrise du sujet comme s'il s'agissait de morceaux de connaissance atomisée*». On a également constaté que, au sein d'une même école, des professeurs poursuivaient des objectifs différents, de sorte que les élèves entraient dans la classe supérieure avec des formations très différentes.

Ainsi, les résultats des études sur l'état actuel des pratiques d'évaluation des professeurs montrent qu'il est nécessaire d'investir de façon très importante dans la formation continue pour bien établir l'évaluation formative. Les changements ne se font pas facilement. Des essais de renforcement de l'évaluation formative dans le curriculum national du Royaume-Uni n'ont eu qu'un succès limité ; ceci est dû en partie au fait que les professeurs interprétaient leur rôle d'évaluateurs uniquement en termes d'évaluation sommative, de sorte que les nouvelles exigences concernant l'évaluation étaient perçues par les professeurs comme une incitation à réaliser leurs propres tests sommatifs (Harlen & Qualter, 1991 ; Scott, 1991). Selon Butler et Beasley (1987), le passage d'une certification externe à une certification propre à l'école, en 1972 au Queensland, a eu peu d'effet sur les pratiques scolaires. Ce n'est qu'après un changement conduit par l'État en 1985, consistant à passer d'une approche sommative à étalonnage normatif à une approche formative centrée sur les critères, que les pratiques ont commencé lentement à changer. Un fait caractéristique du rapport de Butler et Beasley est que les professeurs ont trouvé particulièrement difficile de s'adapter à l'utilisation des résultats sous forme de feedback. Torrance (1986) fait une remarque analogue en ce qui concerne les professeurs en Grande-Bretagne.

Crooks (1988) présente une revue de l'impact sur les élèves des pratiques d'évaluation en classe, revue fondée sur plus de trois cents sources et couvrant de nombreux pays. Il note le contraste entre les nombreux effets positifs d'une bonne évaluation, et le fait que l'évaluation en classe encourage souvent un apprentissage superficiel et par cœur, ainsi qu'une importance exagérée accordée à la notation. Une telle pratique conduit certains élèves à attribuer leurs échecs à leurs faibles capacités, ce qui diminue leur confiance en eux et leur motivation.

Les essais de renforcement de l'évaluation par les professeurs peuvent entraîner une controverse sur les relations entre les fonctions formative et sommative. Alors que l'évaluation formative doit être conduite avec comme objectif principal le feedback vers le processus d'enseignement, elle peut

aussi produire des informations qu'on peut utiliser dans un but sommatif. Ceci peut aider à protéger l'évaluation formative de l'influence écrasante des contrôles externes «à enjeux élevés» (Raizen et al., 1989). Cependant, il peut exister, pour un professeur, une tension entre les rôles de conseiller et de juge ; certains soutiennent que les mêmes outils et méthodes d'évaluation ne peuvent pas remplir deux fonctions différentes, et que l'évaluation ne peut être renforcée qu'à condition de séparer ces fonctions (Harlen et al., 1992). Cependant, ces auteurs ne prennent pas en compte la nécessité d'empêcher le processus sommatif d'envahir le champ du travail de formation (Resnick & Resnick, 1992 ; Black, 1993b).

Là où elles prédominent, les pratiques sommatives peuvent induire en erreur, car les tests externes constituent un mauvais modèle pour l'évaluation formative, et ceci pour les raisons suivantes.

– Dans les tests sommatifs la recherche d'un résultat global unique implique qu'on doit additionner d'une manière arbitraire des données tout à fait disparates (par exemple obtenues sur la partie théorique et sur les travaux pratiques) ; cela ne doit pas être le rôle de l'évaluation formative, qui s'intéresse à la nature des besoins d'apprentissage de chaque élève pris individuellement.

– Le travail sommatif doit insister sur des standards d'uniformité et de validité qui ne sont pas nécessaires dans le travail formatif, qui entravent la liberté et empêchent d'accorder de l'attention aux besoins individuels que demande le travail formatif. En particulier les pratiques formatives n'ont pas à être «justes»; leur priorité est l'identification des besoins de chacun des élèves, ce qui implique des traitements différenciés pour des élèves différents.

– Un travail à but sommatif peut exiger la collecte de documents justificatifs adéquats pour les résultats, ce qui augmente la charge de travail, tandis que le travail formatif demande d'agir à partir des données plutôt que de les stocker.

## ÉVALUATION ET APPRENTISSAGE

Le principe clé que l'on veut souligner ici est que l'évaluation formative doit être intimement liée aux processus d'enseignement et d'apprentissage. Ce point a été clairement exposé par Linn (1989) :

*«La construction de tests pour la prise de décisions éducatives dans la classe exige une intégration des tests et de l'enseignement. Elle exige aussi une conception claire du curriculum, des objectifs et des méthodes d'enseignement. Elle demande enfin une théorie de*

*l'enseignement et de l'apprentissage ainsi qu'une bien meilleure compréhension des processus cognitifs des apprenants.»*

Cette nécessité pour les évaluations de servir de base à un enseignement dirigé vers l'amélioration de la pensée a été explorée par Resnick et Resnick (1992). Les tests composés de questions courtes «atomisées» semblent conforter des hypothèses sur l'apprentissage que la psychologie actuelle a abandonnées. Brown et al. (1992) soutiennent cette position en s'appuyant sur des résultats qui montrent que l'amélioration des méthodes diminue la corrélation entre les résultats et les mesures initiales du QI des élèves. Citons par exemple l'hypothèse fautive qu'on peut enseigner une capacité complexe en la décomposant en parties enseignées et évaluées séparément. On peut citer également l'hypothèse fautive selon laquelle une idée commune à plusieurs situations peut être enseignée plus économiquement en la présentant dans toute son abstraction, pour qu'elle puisse être utilisée ensuite dans de nombreuses situations.

Si le feedback apporté par des évaluations fréquentes doit améliorer l'apprentissage, alors il doit en refléter les buts importants. Berryman (1990) remarque que le commerce et l'industrie accordent la priorité à l'autonomie, à la capacité de travailler en groupe et à résoudre des conflits, ainsi qu'au développement de capacités intellectuelles d'ordre élevé. Newmann et Archbald (1992) soutiennent que «*la réussite universitaire authentique*» demande une compréhension en profondeur, et des capacités d'intégration, de collaboration et d'échanges approfondis. La prise en compte de tels buts est pratiquement inexistante dans la plupart des tests externes ou standardisés.

Si l'information provenant de l'évaluation doit avoir un sens par rapport aux objectifs d'apprentissage du curriculum, alors elle doit être de type critique. Le développement de tests critiques s'est révélé très compliqué. Le plus souvent, la difficulté provient du fait que si les critères sont généraux et vagues, leur valeur formative risque d'être perdue, alors que s'ils sont trop spécifiques, les professeurs vont se noyer dans leur multiplicité, tandis que l'enseignement sera fragmenté de telle sorte que les aspects holistiques et relationnels seront perdus (Popham, 1987).

Si l'évaluation doit guider l'apprentissage pendant le déroulement d'un programme, alors elle doit refléter les critères qui sont formulés dans une séquence d'apprentissage. En principe, ceci peut constituer un guide très précieux pour suivre les progrès d'un individu et localiser les obstacles à son apprentissage. Cependant, on ne peut le réaliser sans une connaissance réellement fondée des séquences d'apprentissage des élèves (Simon et al., 1994). Des tentatives peu rigoureuses conduisent à des critères flous et inopérants, comme «*comprendre parfaitement*» suivi de «*comprendre plus parfaitement*».

## VALIDITÉ ET FIABILITÉ

Les tests externes sous forme écrite bénéficient de la part du public d'une confiance plus grande qu'ils ne devraient. Les limites de leur validité et de leur fiabilité ne sont pas bien appréciées. Toute tentative d'amélioration de la qualité d'un mode d'évaluation doit s'appuyer sur un concept général de validité qui soit convaincant. La première phrase de la revue de Messick (1989) donne une définition qui fait autorité :

*«La validité est un jugement évaluatif intégré sur le degré de confirmation, par des données expérimentales et des considérations théoriques, de la justesse et de l'opportunité de conclusions et d'actions fondées sur les résultats d'examens ou d'autres tests d'évaluation.»*

Le domaine de ces *conclusions* et de ces *actions* a été étendu lors d'études récentes. Par exemple, Moss (1992) affirme que l'utilisation qu'on en fera et l'effet de reflux sur l'enseignement sont deux aspects à prendre en compte dans l'estimation de la validité. Ainsi, si une méthode d'évaluation paraissait renforcer des styles d'enseignement qui ne sont pas en accord avec les objectifs de l'éducation, sa validité pourrait être mise en doute sur cette seule base. Plus une activité d'évaluation se rapproche de l'activité réelle pour laquelle ses résultats doivent être considérés comme pertinents, plus elle a de chances de satisfaire au critère de validité. Vue sous cet angle, l'évaluation faite en classe a de meilleures chances de succès que les tests écrits formels et en temps limité.

La fiabilité des évaluations en classe est par nature difficile à explorer par suite de l'absence de toute mesure indépendante des acquisitions «réelles» des élèves. Hoge et Coladarci (1989) proposent une revue exhaustive de la littérature concernant les jugements portés par les professeurs sur les résultats scolaires. Ils concluent que les appréciations internes peuvent avoir une validité supérieure à celle des tests psychologiques standard correspondants.

La fiabilité limitée des tests formels n'est pas prise en compte et, dans beaucoup de systèmes nationaux, n'est pas mesurée. La cohérence interne est évidemment prisée, mais pour des raisons diverses, les élèves ont dans l'ensemble de moins bons résultats aux tests formels, comme le démontre la revue de Nuttall (1987). Pour un certain nombre de tests en science, Gauld (1980) constate que les élèves se trompent souvent sur la question posée, paraissent incompetents à cause d'une seule erreur dans un processus complexe, ne réussissent pas à utiliser ce qu'ils savent parce qu'ils le croient sans relation avec la question, et peuvent obtenir des notes trop basses parce que le correcteur n'a pas compris le raisonnement conduisant à une réponse inattendue.

En tout état de cause, pour l'évaluation formative, les critères de validité et de fiabilité doivent être reformulés. La validité de l'évaluation formative doit être appréciée en relation avec son efficacité dans l'amélioration de l'enseignement.

## **AMÉLIORATION DE L'ÉVALUATION FORMATIVE LES BESOINS DES ENSEIGNANTS**

La technologie de collecte des données sur les progrès des élèves commence seulement à se développer. La plupart des enseignants ont toujours utilisé de manière informelle une grande variété de sources, mais ceci peut induire en erreur. Très souvent par exemple, un enseignant utilise, en dialoguant avec la classe, les réponses de quelques élèves seulement, pour se rassurer sur le déroulement du plan de travail. Il est nécessaire d'affiner les pratiques pour obtenir des résultats plus complets. Les livres de Fairbrother et al. (1993, 1995) donnent des exemples de fiches de travail pour l'auto-évaluation, qui formulent des objectifs d'apprentissage pour le travail en classe ; pour chacun de ces objectifs, l'élève doit consigner si, selon lui, il l'a atteint ou non. Ces fiches donnent au professeur l'information critériée essentielle pour l'évaluation formative. De plus, parce qu'elles fournissent de manière systématique des informations écrites, elles déchargent le professeur d'avoir à noter et enregistrer les résultats à partir des seules observations éphémères des événements de la classe. L'observation par les enseignants peut être irremplaçable : certains ont trouvé étonnamment utile de suspendre pour un temps leurs interventions pédagogiques actives – en expliquant bien à la classe ce qu'ils font et pourquoi – et de se concentrer uniquement sur l'observation et l'écoute d'un petit nombre d'élèves (cf. Cavendish et al., 1990 ; Connor, 1991).

Cependant, l'enjeu ici dépasse largement la collecte de données sur les «erreurs». Comme l'a montré Bonniol (1991), l'enseignant a besoin d'en comprendre les causes, ce qui mettrait à l'épreuve le meilleur expert en apprentissage. De plus, l'évaluation formative ne peut pas être simplement ajoutée aux schémas de travail existants ; elle doit être incorporée dans ces schémas, ne serait-ce que parce que son utilisation pour le guidage des élèves, selon les besoins de chacun, ne peut avoir lieu que si l'enseignement est organisé de façon suffisamment souple. C'est l'aspect le plus exigeant du travail formatif.

Une manière «macroscopique» et à long terme de traiter les besoins différents est de répartir les élèves dans différentes classes, en fonction de leurs performances antérieures. C'est une solution brutale qui ne peut pas répondre à des besoins variés et à court terme. Des réponses plus fines peuvent être apportées par l'organisation des cours en modules. Pendant

les deux tiers de la durée du module on emploie des méthodes d'évaluation informelle et continue, puis une évaluation plus formelle est utilisée pour déterminer le travail d'apprentissage différencié qui se fera dans le dernier tiers ; ceux qui ont maîtrisé les idées essentielles poursuivent par un travail d'extension, tandis que ceux qui ont trouvé des difficultés particulières travaillent sur un matériel destiné à les aider à franchir ces obstacles. Ainsi certains réussiront dans un domaine plus étendu que d'autres, mais on peut espérer que tous auront compris le minimum nécessaire pour donner un sens à leur travail ultérieur (cf. Black H., 1993 ; Hunt, 1993). Il existe d'autres approches moins formelles et plus souples, comportant des occasions de révisions ou de reprises pour ceux qui en ont besoin.

Ce dernier point, et beaucoup d'autres, sont en relation avec la plus ou moins grande souplesse des programmes d'enseignement. Là où un travail ouvert, travail expérimental ou projet avec des recherches bibliographiques, trouve place dans l'enseignement scientifique, une attention particulière à la sélection et à la définition des tâches peut permettre une bonne différenciation fondée sur l'ajustement de la tâche aux possibilités et aux progrès de l'élève. Cependant il apparaît que ce sont les meilleurs élèves qui bénéficient d'un cadre ouvert et souple. Ceux dont les progrès sont plus lents manquent de la confiance et de la vision claire de leurs besoins qui sont indispensables pour être autonomes et efficaces. Ainsi, pour eux, un cadre plus rigide peut être nécessaire dans un travail ouvert.

L'utilisation efficace du feedback de l'évaluation exige de l'enseignant qu'il exerce son jugement, et puisse conduire son programme d'enseignement avec la confiance et la souplesse qui ne peuvent venir que de l'appropriation de ce programme. Ainsi il semble que, idéalement, tout schéma permettant d'incorporer de bonnes occasions d'évaluation formative doit être construit par les professeurs eux-mêmes pour leurs besoins propres. Le besoin d'une formation continue approfondie pour soutenir cet aspect de l'enseignement est illustré par le rapport de Torrie (1989) sur un travail en formation continue avec un groupe d'enseignants australiens. Ceux-ci ont produit puis dégraissé une longue liste constituée de leurs propres critères de performances, et les ont ensuite inscrits dans une grille de progression d'apprentissages. Puis la liste a été encore affinée lorsqu'ils ont essayé de produire les exercices d'évaluation correspondants. Les enseignants en ont tiré, outre une plus grande confiance dans leur enseignement, l'opinion que leur travail d'évaluation antérieur avait été médiocre car ils n'avaient pas vraiment évalué les objectifs de leur enseignement. Des conclusions semblables ont été présentées dans une revue des développements dans plusieurs états australiens par Withers (1987). Par ailleurs, pour assurer une cohérence dans l'aide aux élèves, les enseignants doivent mettre en commun leurs informations et leurs pratiques d'évaluation. L'échange de travaux d'élèves est valable en soi comme base

pour établir une collaboration entre les responsables. Stage (1990), Baird et al. (1991) et Wood (1991) décrivent tous l'utilisation d'une «réflexion collective» dans la formation des enseignants.

Incorporer l'évaluation formative dans leur enseignement engage les enseignants bien au-delà de l'acquisition des compétences strictement nécessaires ; elle implique aussi de profonds changements dans leur rôle (Tobin et al., 1988). En particulier, puisque beaucoup d'élèves peuvent avoir pris l'habitude d'en faire juste assez pour s'en tirer, ou bien ont cessé de croire qu'ils pourraient être compétents, le contrat entre l'enseignant et l'élève doit être reformulé. Transférer plus de responsabilités aux élèves est une manière de réduire le poids de la tâche ; il existe d'autres raisons plus importantes d'effectuer ce transfert, qui vont être discutées ci-dessous.

## **AMÉLIORATION DE L'ÉVALUATION FORMATIVE LES BESOINS DES ÉLÈVES**

Dans leurs revues sur les principes de l'évaluation formative, Perrenoud (1991) et Bonniol (1991) insistent sur le fait que le développement de ce type d'évaluation implique des modifications du rôle des enseignants aussi bien que des élèves. Les élèves doivent comprendre le cadre de référence de l'enseignant et le modèle d'apprentissage qui donne du sens aux critères qui sont reflétés dans l'évaluation. C'est sur cette compréhension que peut s'établir la confiance des élèves en leurs facultés, liée à l'assurance qu'on pourra faire quelque chose pour leurs difficultés.

Un manque de compréhension entre élèves et enseignants peut affaiblir toute évaluation informelle – comme l'illustre un article de Perrin (1991) sur l'étude d'élèves de l'école primaire dans le canton de Genève : ces élèves croyaient que les évaluations étaient faites au profit de l'école et de leurs parents, et non pas au leur. Les élèves faibles pensaient que le but était de les faire travailler plus. Comme l'évaluation n'était pas mise à profit pour leur dire comment travailler autrement, ils la voyaient comme une source de tension qui les rendait anxieux. Suite à ces constatations, le canton de Genève a décidé de réduire les tests sommatifs et d'augmenter le rôle formatif de l'évaluation. La charge de travail donnée aux élèves par de fréquents contrôles sommatifs peut ne pas être perçue de cette manière s'il existe un feedback et des occasions d'améliorer leurs performances, comme l'a montré l'étude d'Iredale (1990) sur l'attitude des élèves envers un mode d'évaluation avec notation, en sciences.

Les élèves ne peuvent prendre une part effective dans leur propre évaluation que dans le cadre d'un programme à long terme, conçu pour les aider à atteindre et maintenir une vue globale des objectifs de leur

apprentissage, ainsi qu'à appliquer les critères d'évaluation à leurs propres progrès. Les élèves n'arriveront à cette vue globale que si ceci est explicitement enseigné. Les rapports de Baird et Mitchell (1986), Baird et Northfield (1992) et Fairbrother et al. (1993, 1995) montrent qu'un tel enseignement demande du temps avant d'aboutir à des succès. Il comporte la traduction des objectifs du curriculum dans un langage que tous les élèves peuvent comprendre, et ce jusqu'à un niveau de détail qui les aide à faire le lien entre ces objectifs et leurs efforts d'apprentissage. D'autre part, les cibles doivent pouvoir être atteintes à court terme, et être raisonnablement modestes par rapport aux chances de succès de l'apprenant.

La capacité de jugement des élèves sur leur propre travail est d'une importance capitale dans l'apprentissage. Tobin et al. (1988), Labudde et al. (1988) ont tous insisté sur le rôle clé de l'auto-évaluation. Zessoules et Gardner (1991), en décrivant le Harvard Project Zero, soulignent que l'auto-évaluation pendant l'apprentissage est une composante cruciale pour le développement de la compréhension de sujets complexes, appuyée sur des habitudes mentales de réflexion intellectuelle ; c'est pourquoi leur projet vise à réaliser des conditions permettant de s'assurer que les élèves sont actifs et réfléchis au cours de l'auto-évaluation. D'autres ont soutenu que la méta-cognition – par quoi ils entendent prise de conscience et autonomie quant à la nature des tâches d'apprentissage – est essentielle pour que les élèves progressent dans l'apprentissage des concepts (cf. Brown, 1987 ; White & Gunstone, 1989).

Il est clair que c'est grâce à l'implication des élèves que les enseignants pourront plus facilement mener à bien un programme d'évaluation formative. Cependant, cette implication change aussi bien le rôle de l'élève comme apprenant que la nature des relations entre l'enseignant et l'élève, en faisant porter à ce dernier une plus grande part de responsabilité dans l'apprentissage. Ainsi l'amélioration de l'évaluation formative peut conduire à des changements très significatifs – changements qui devraient aider puissamment au développement personnel des élèves, et qui devraient faire partie de tout programme visant à en faire des apprenants plus efficaces.

## CONCLUSIONS

Si les enseignants sont disposés à tenter des améliorations, un organisme de formation peut fournir un forum en terrain neutre où ils pourront recevoir les conseils et les idées que peut offrir la littérature, et échanger des expériences avec d'autres personnes travaillant avec les mêmes objectifs. Les thèmes abordés dans cet article pourraient fournir un cadre pour bâtir

en partie un tel programme. Cependant, ce qui est essentiel est le travail fait par les enseignants pour mettre en place des changements dans leurs propres classes, accompagné de réunions régulières de suivi où les succès comme les échecs pourront être rapportés et comparés.

L'ensemble de la littérature de recherche est couvert dans les livres de Wood (1991) et Gipps (1994) et, plus particulièrement pour les sciences, dans la revue de Black (1993b). On peut trouver des conseils généraux sur les activités pratiques en classe dans les nombreux articles cités, en particulier ceux de Baird et Northfield (1992) et Fairbrother et al. (1995). Parmi les nombreux livres généraux sur l'évaluation parus récemment, ceux d'Airasian (1991) et de Stiggins (1994) sont particulièrement centrés sur l'évaluation formative.

Dans leurs responsabilités concernant l'évaluation, les enseignants doivent concilier les besoins d'apprentissage de leurs élèves, dont la satisfaction doit être leur souci premier, avec l'obligation d'obtenir de bons résultats aux tests nationaux réglementaires, ainsi qu'avec le respect du cadre des règlements de l'école et les attentes des parents. Ces exigences sont souvent contradictoires. S'ils veulent être en position de réaliser les changements radicaux qui sont réclamés dans cet article, les enseignants auront besoin de travailler ensemble pour s'épauler face à ces pressions, et de faire appel à l'aide et à la compréhension qui pourront être offertes par ceux qui sont à l'extérieur de l'école.

Finalement, l'analyse de l'influence du contexte systémique des tests et de l'évaluation sur les pratiques des enseignants montre qu'une action au niveau du public et du politique est essentielle, afin d'établir et soutenir à la fois les contextes systémiques et l'approbation du public qui sont nécessaires pour la survie de telles innovations.

## BIBLIOGRAPHIE

- AIRASIAN P.W. (1991). *Classroom Assessment*. New York, McGraw Hill.
- BAIRD J.R. & MITCHELL I.J. (Eds) (1986). *Improving the quality of teaching and learning. An Australian case study - the PEEL project*. Melbourne, Monash University.
- BAIRD J.R. & NORTHFIELD J.R. (Eds) (1992). *Learning from the PEEL Experience*. Melbourne, Monash University.
- BAIRD J.R., FENSHAM P.J., GUNSTONE R.F. & WHITE R.T. (1991). The importance of reflection in improving science teaching and learning. *Journal of Research in Science Teaching*, vol. 28, n° 2, pp. 163-182.
- BERRYMAN S.E. (1990). Sending clear signals to schools and labour markets. In J.L. Schwartz & K.L. Viator (Eds), *The Prices of Secrecy : the Social, Intellectual and Psychological Costs of Secrecy*. Cambridge Massachussets, ETC Harvard Graduate School of Education, pp. 35-45.

- BLACK H. (1986). Assessment for learning. In D.L. Nuttall (Ed.), *Assessing Educational Achievement*. London, Falmer Press, pp. 7-18.
- BLACK H. (1993). Assessment : A Scottish Model. In R. Fairbrother, P.J. Black & P. Gill (Eds), *TAPAS : Teacher Assessment of Pupils : Active Support*. King's Education Papers n°3. London, Centre for Educational Studies King's College, pp. 91-94.
- BLACK P.J. (1993a). Formative and Summative Assessment by Teachers. *Studies in Science Education*, n° 21, pp. 49-97.
- BLACK P.J. (1993b). Assessment policy and public confidence : Comments on the BERA Policy Task Group's article «Assessment and the improvement of education». *The Curriculum Journal*, vol. 4, n° 3, pp. 421-427.
- BONNIOL J.-J. (1991). The mechanisms regulating the learning process of pupils : contribution to a theory of formative assessment. In P. Weston (Ed.), *Assessment of Pupils Achievement : Motivation and School Success*. Amsterdam, Swets and Zeitlinger, pp. 119-137.
- BROWN A. (1987). Metacognition, executive control, self-regulation and other more mysterious mechanisms. In F.E. Weinert & R.H. Kluwe, *Metacognition, Motivation and Understanding*. Hillsdale NJ, Lawrence Erlbaum, pp. 65-116.
- BROWN A.L., CAMPIONE J.C., WEBBER L.S. & Mc GILLY K. (1992). Interactive Learning Environments : A New Look at Assessment and Instruction. In B.R. Gifford & M.C. O'Connor (Eds), *Changing Assessments : Alternative Views of Aptitude, Achievement and Instruction*. Boston USA & Dordrecht Netherlands, Kluwer, pp. 121-211.
- BUTLER J.E., BEASLEY W.F., BUCKLEY D. & ENDEAN L. (1980). Pupil task involvement in secondary classrooms. *Research in Science Education*, n° 10, pp. 93-106.
- BUTLER J.E. & BEASLEY W.F. (1987). The impact of assessment changes on the science curriculum. *Research in Science Education*, n° 17, pp. 236-243.
- CAVENDISH S., GALTON M., HARGREAVES L. & HARLEN W. (1990). *Observing Activities*. London, Paul Chapman.
- CONNOR C. (1991). *Assessment and Testing in the Primary School*. London, Falmer Press.
- CROOKS T.J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, vol. 58 , n° 4, pp. 438-481.
- DES (1987). *National curriculum : Task Group on Assessment and Testing : A Report*. London, Department of Education and Science & Welsh Office.
- FAIRBROTHER R.W., BLACK P.J. & GILL P. (Eds) (1993). *TAPAS : Teacher Assessment of Pupils : Active support for Teachers*. King's Education Papers n° 3. London, Centre for Educational Studies King's College.
- FAIRBROTHER R.W., BLACK P.J. & GILL P. (Eds) (1995). *Teachers Assessing Pupils : Lessons from Science Classrooms*. Hatfield, Association for Science Education.
- GAULD C.F. (1980). Subject oriented test construction. *Research in Science Education*, n° 10, pp. 77-82.
- GIPPS C.V. (1994). *Beyond Testing : Towards a Theory of Educational Assessment*. London, Falmer Press.
- GRISAY A. (1991). Improving assessment in primary schools : «APER» research reduces failure rates. In P. Weston (Ed.), *Assessment of Pupils Achievement : Motivation and School Success*. Amsterdam, Swets & Zeitlinger, pp. 103-118.
- HANEY W. (1991). We Must Take Care : Fitting Assessments to Functions. In V. Perrone (Ed.), *Expanding Student Assessment*. Alexandria, Virginia, Association for Supervision and Curriculum Development, pp. 42-163.
- HARLEN W. & QUALTER A. (1991). Issues in SAT development and the practice of teacher assessment. *Cambridge Journal of Education*, vol. 21, n° 2, pp. 141-152.

- HARLEN W., GIPPS C., BROADFOOT P. & NUTTALL D. (1992). Assessment and the improvement of education. *The Curriculum Journal*, vol. 3, n° 3, pp. 215-230.
- HODSON D. (1986). The role of assessment in the «Curriculum Cycle» : a survey of science department practice. *Research in Science and Technological Education*, vol. 4, n° 1, pp. 7-17.
- HOGUE R.D. & COLADARCI T. (1989). Teacher-based judgments of academic achievement : a review of literature. *Review of Educational Research*, vol. 59, n° 3, pp. 297-313.
- HUNT A. (1993). *Pathways through Science Nuffield Modular Course*. Longmans, Harlow UK.
- IREDALE C. (1990). Pupils' attitudes towards GASP (Graded Assessments in Science Project). *School Science Review*, vol. 72, n° 258, pp. 133-137.
- LABUDDE P., REIF F. & QUINN L. (1988). Facilitation of scientific concept learning by interpretation procedures and diagnosis. *International Journal of Science Education*, vol. 10, n° 1, pp. 81-98.
- LINN R.L. (1989). Current Perspectives and Future Directions. In R.L. Linn (Ed.), *Educational Measurement (3rd Edition)*. London, Collier Macmillan, pp. 1-10.
- MESSICK S. (1989). Validity. In R.L. Linn (Ed.), *Educational Measurement (3rd Edition)*. London, Collier Macmillan, pp. 12-103.
- MOSS P.A. (1992). Shifting conceptions of validity in educational measurement : implications for performance assessment. *Review of Educational Research*, vol. 62, n° 3, pp. 229-258.
- NEWMANN F.M. & ARCHBALD D.A. (1992). The Nature of Authentic Academic Achievement. In Berlak et al. (Eds.), *Toward a New Science of Educational Testing and Assessment*. Albany, State University of New York Press, pp. 71-83.
- NUTTALL D.L. (1987). The validity of assessments. *European Journal of Psychology of Education*, vol. II, n° 2, pp. 109-118.
- PERRENOUD P. (1991). Towards a pragmatic approach to formative evaluation. In P. Weston (Ed.), *Assessment of Pupils Achievement : Motivation and School Success*. Amsterdam, Swets & Zeitlinger, pp. 79-101.
- PERRIN M. (1991). Summative evaluation and pupil motivation. In P. Weston (Ed.), *Assessment of Pupils Achievement : Motivation and School Success*. Amsterdam, Swets & Zeitlinger, pp. 169-173.
- POPHAM W.J. (1987). Two-plus decades of educational research. *International Journal of Educational Research*, vol. 11, n° 1, pp. 31-41.
- RAIZEN S.A., BARON J.B., CHAMPAGNE A.B., HAERTEL E., MULLIS I.V.S. & OAKES J. (1989). *Assessment in Elementary School Science Education*. Washington, National Centre for Improving Science Education.
- RESNICK L.B. & RESNICK D.P. (1992). Assessing the Thinking Curriculum : New Tools for Educational Reform. In B.R. Gifford & M.C. O'Connor (Eds), *Changing Assessments : Alternative Views of Aptitude, Achievement and Instruction*. Boston et Dordrecht, Kluwer, pp. 37-75.
- RUDMAN H.C. (1987). Testing and teaching : two sides of the same coin ? *Studies in Educational Evaluation*, n° 13, pp. 73-90.
- SCOTT D. (1991). Issues and themes : coursework and coursework assessment in the GCSE. *Research Papers in Education*, vol. 6, n° 1, pp. 3-19.
- SIMON S., BLACK P.J., BROWN M. & BLONDEL E. (1994). Progression in Understanding the Equilibrium of Forces. *Research Papers in Education*, vol. 9, n° 2, pp. 249-280.
- STAGE E.K. (1990). The Psychological Costs of Secrecy and the Promise of Openness : the Impact on the Teacher. In J.L. Schwartz & K.L. Viator (Eds), *The Prices of Secrecy : the Social, Intellectual and Psychological Costs of Secrecy*. Cambridge, Massachusetts, ETC Harvard Graduate School of Education, pp. 93-98.

- STIGGINS R.J. (1994). *Student-Centered Classroom Assessment*. New York, Merrill Macmillan.
- TOBIN K., CAPIE W. & BETTENCOURT A. (1988). Active teaching for higher cognitive learning in science. *International Journal of Science Education*, vol. 10, n° 1, pp. 17-27.
- TOBIN K., ESPINET M., BYRD S.E. & ADAMS A. (1988). Perspectives of effective science learning. *Science Education*, vol. 72, n° 4, pp. 433-451.
- TORRANCE H. (1986). Expanding school based assessment : issues, problems and future possibilities. *Research Papers in Education*, vol. 1, n° 2, pp. 48-59.
- TORRIE I. (1989). Developing achievement based assessment using grade related criteria. *Research in Science Education*, n° 19, pp. 286-290.
- WHITE R.T. & GUNSTONE R.F. (1989). Meta-learning and conceptual change. *International Journal of Science Education*, vol. 11, pp. 577-586.
- WITHERS G. (1987). From marking strategy to assessment procedure : a review of recent Australian practices. *Studies in Educational Evaluation*, n° 13, pp. 7-19.
- WOOD R. (1991). *Assessment and Testing : a Survey of Research*. Cambridge, Cambridge University Press.
- ZESSOULES R. & GARDNER H. (1991). Authentic Assessment : Beyond the Buzzword and Into the Classroom. In V. Perrone (Ed.), *Expanding Student Assessment*. Alexandria, Virginia, Association for Supervision and Curriculum Development, pp. 47-71.