

Revue trimestrielle
n° 244 décembre 1996

EDUCATION

formation

L'ÉVALUATION SOMMATIVE DANS L'ENSEIGNEMENT DES SCIENCES

Babacar GUEYE (**)

(*) Maître-Assistant à l'Ecole Normale Supérieure, B.P. 5036
Dakar-Sénégal

Résumé

La question de l'évaluation a depuis très longtemps occupé une place importante dans les recherches en Sciences de l'éducation et a donné naissance aux différentes taxinomies que nous connaissons. Cependant, force est de reconnaître que dans le cas particulier de l'évaluation sommative en Sciences le manque de validité et de fidélité des différentes épreuves proposées dans l'enseignement secondaire est un problème qui interpelle les didacticiens.

Mots clés

Évaluation, curriculum, validité, fidélité, Sciences.

Introduction

Qu'elle soit consciente ou non, l'évaluation est toujours présente dans l'enseignement (interrogations orales et écrites, examens, concours, etc.); c'est ce qui sans aucun doute explique la place centrale qu'on lui accorde dans tout curriculum bien construit.

Cependant il est généralement admis que de tous les éléments d'un curriculum, l'évaluation est le plus difficile à mettre en place; de telle sorte que dès leur émergence

- recueillir un ensemble d'informations suffisamment pertinentes, valides et fiables;
- examiner le degré d'adéquation entre cet ensemble d'informations et un ensemble de critères adéquats aux objectifs fixés au départ ou ajustés en cours de route;
- en vue de prendre une décision (DE KETELE 1989).

les Sciences de l'éducation en ont fait un domaine privilégié d'investigation.

Ainsi naquit la docimologie, discipline qui étudie les pratiques d'évaluation pour montrer les bons fonctionnements et les dysfonctionnements afin de consolider les premiers et de remédier aux seconds.

C'est dans ce champ de la docimologie critique que nous allons nous placer dans le cas particulier des Sciences.

Après avoir clarifié le concept d'évaluation dans son acception scolaire ainsi que sa problématique, nous examinerons la qualité des instruments d'évaluation sommative en Sciences à l'école secondaire et leurs conséquences sur les autres éléments du curriculum scientifique afin de dégager quelques perspectives possibles d'amélioration.

1. Définitions du concept d'évaluation

Au fil de l'histoire de l'éducation, de nombreux auteurs ont essayé de définir le concept d'évaluation avec plus ou moins de succès et de pertinence.

Aujourd'hui nous pouvons constater que la plupart des auteurs mettent l'évaluation au service de la prise de décision (STUFFLEBEAM 1980, DE KETELE 1989).

Par exemple, DE KETELE définit l'évaluation en ces termes. Évaluer signifie:

Cette nouvelle définition du concept d'évaluation semble correspondre à une rupture sur le plan épistémologique puisque si aujourd'hui la prise de décision en toute connaissance de cause semble être le but déclaré de l'évaluation, il fut un temps assez long pendant lequel le seul désir primitif et subjectif par essence de porter un jugement de valeur à partir de mesures semblait

l'emporter sur tout le reste. D'ailleurs en fin de compte, on n'hésitait plus à considérer l'évaluation comme étant un jugement de mérite, parfois fondé exclusivement sur la mesure, comme les données fournies par le rendement aux tests, mais qui implique le plus souvent la synthèse de plusieurs mesures, d'incidents critiques, d'impressions toutes subjectives et d'autres sortes d'évidences (EBEL, 1965).

En réalité, jugement de valeur et prise de décision sont les deux étapes d'un même processus.

Toute prise de décision s'effectue dans tous les cas à partir d'un jugement sur la valeur des actions ou sur la performance des individus par rapport à des objectifs, implicites ou explicites.

D'ailleurs une autre définition de l'Évaluation la considère comme devant servir à déterminer la congruence entre la performance et les objectifs. C'est celle donnée par TYLER en ces termes:

« Le processus de l'évaluation consiste essentiellement à déterminer dans quelle mesure les objectifs d'éducation sont en voie d'être atteints... cependant comme les objectifs d'éducation visent essentiellement à changer les êtres humains c'est à dire que le but est de produire dans le comportement des étudiants certains changements

souhaitables alors que l'évaluation est le processus consistant à déterminer dans quelle mesure ces changements de comportements sont en train de se produire » (TYLER, 1950).

2. Problématique de l'évaluation en Sciences

Toutes les disciplines scolaires ont tenté de répondre aux questions suivantes à propos de leur évaluation:

- Qu'évaluer ?
- Quand et pourquoi évaluer ?
- Comment évaluer ?

1. Qu'évaluer ?

La réponse traditionnellement donnée par l'école à cette question découle sans doute de la conception suivante à propos de l'élaboration d'un programme qui doit se faire en quatre phases essentielles :

- Déterminer les objectifs que le cours ou le programme devrait viser à atteindre.

- Choisir les expériences d'apprentissage qui aideront à atteindre ces objectifs.
- Organiser ces expériences d'apprentissage.
- Déterminer dans quelle mesure on a atteint les objectifs (FURST, 194).

Il s'agit donc d'évaluer les objectifs, plus précisément les objectifs en terme de comportements si nous rapprochons les définitions de TYLER (1950) et de FURST (1964).

Dès lors, la tâche essentielle était de repérer ces objectifs et de les catégoriser. Évidemment, pour les raisons que nous avons précédemment évoquées, c'est l'approche comportementaliste qui a été privilégiée.

Le résultat de tout cela a été l'émergence des célèbres taxinomies que nous connaissons (BLOOM, D'HAINAUT, etc.) et qui ont essayé chacune de répertorier, de décrire et de classer ce que l'élève doit pouvoir faire indifféremment du contenu matière qui sert de support.

La première de ces taxinomies, celle de BLOOM publiée en 1956 qui comprend six niveaux (connaissance, compréhension, application, analyse, synthèse, évaluation) fut utilisée au Québec dès 1964, puis en Belgique en 1972 dans l'évaluation des élèves.

Une telle utilisation avait (sans doute) l'avantage de rationaliser, systématiser, évaluer une action éducative trop longtemps abandonnée à l'intuition, à la sensibilité, au bon sens (DE LANDSHEERE, 1975). Cependant elle a fait l'objet de nombreuses critiques que nous pouvons résumer ainsi: l'usage des taxinomies entraîne une

concentration trop excessive de l'évaluation et de l'enseignement sur des comportements atomisés qui ne tiennent nullement compte de l'initiative et du désir de l'apprenant réduit à une quasi inactivité.

C'est face à ce problème réel et au fait qu'au niveau de l'enseignement des Sciences au secondaire on mettait l'accent sur le processus de la découverte scientifique qu'un autre courant un peu moins généraliste et moins célèbre s'est développé à propos des Sciences. Là on a essayé de rapprocher les comportements de la spécificité de la matière elle-même en définissant des critères de performance par rapport à des capacités et des connaissances relatives aux sciences.

Il s'agit là d'une avancée assez significative qui concerne aussi bien les objectifs cognitifs que les objectifs affectifs et qu'illustrent les travaux de MUNRO (1975).

2. Quand et pourquoi évaluer?

Il est de tradition à l'école d'évaluer en cours d'apprentissage (interrogations orales ou écrites, compositions, etc.) ou en fin d'apprentissage (examen, concours). SCRIVEN (1967) distingue ces deux moments en évaluation formative pour la première et évaluation sommative pour la seconde.

Ces deux types d'évaluation se distinguent non seulement par le moment où elles ont lieu mais aussi par la fonction qui est la leur dans le fonctionnement des programmes d'enseignement.

L'évaluation formative peut se définir ainsi : un processus d'évaluation continue ayant pour objet d'assurer la progression de chaque individu dans une démarche d'apprentissage avec l'intention de modifier la situation d'apprentissage ou le rythme de cette progression pour apporter (s'il y a lieu) des améliorations ou des correctifs appropriés (SCALLON, 1988).

L'évaluation sommative qu'on a de plus en plus tendance à appeler évaluation certificative est définie comme celle devant déboucher sur une décision dichotomique de réussite ou d'échec relatif à une période d'apprentissage, d'acceptation ou de rejet d'une promotion, de poursuite d'une action ou de l'arrêt de celle-ci (DE KETELE et ROEGIERS, 1993). Ce qu'il faut noter ici, c'est que selon le moment où l'on évalue à l'école, les personnes qui évaluent, les moyens mis en oeuvre, l'environnement, la psychologie des formateurs et des formés varient énormément.

Dans le cas particulier des Sciences, il semble qu'en l'état actuel l'évaluation sommative pose plus de problèmes que l'évaluation formative.

En effet, le «centralisme pseudo-démocratique» adopté pour effectuer de manière transparente et équitable l'évaluation des objectifs empêche lors de l'évaluation certificative, du fait du nombre de candidats et de la durée limitée de l'épreuve, la mobilisation de toutes les ressources didactiques dont disposent les disciplines scientifiques dans leur enseignement quotidien au niveau de chaque établissement scolaire.

3. Comment évaluer?

Dans la plupart des pays ce sont les professeurs qui enseignent la discipline, qui élaborent les épreuves d'évaluation formative et sommative et l'apprenant doit faire ses preuves à partir d'une production écrite : c'est la fameuse épreuve « papier crayon » commune à toutes les disciplines scolaires. Certes l'épreuve orale est pratiquée mais c'est l'écrit qui a toujours le plus fort coefficient.

L'école utilise essentiellement deux techniques d'évaluation à l'écrit :

- Le test dit objectif dans lequel l'élève doit opérer un choix parmi plusieurs réponses dont une seule est juste. C'est ce genre d'épreuves qui a résulté des premières applications des taxinomies notamment au Canada en 1964 et en Belgique en 1972 en ce qui concerne la taxinomie de BLOOM.
- L'épreuve à réponse rédigée : ici l'élève élabore sa réponse, l'organise et la présente à son gré avec une certaine marge de liberté.

En Sciences, l'épreuve à réponse rédigée a évolué d'une forme traditionnelle en une question, une phrase ou un seul mot à une forme plus structurée en plusieurs questions qui appellent des réponses plus ou moins complémentaires.

Cette évolution s'est faite en même temps qu'on introduisait dans l'enseignement scientifique tout le processus méthodologique qui permet de faire des découvertes et de comprendre les phénomènes.

3. Approche critique de l'évaluation en Sciences

Pour les raisons évoquées plus haut, cette étude critique portera sur l'évaluation sommative en Sciences.

Il s'agit essentiellement de voir dans quelle mesure les révélateurs proposés ici et là permettent aux apprenants de manifester les qualités cognitives et affectives que l'enseignement scientifique s'efforce d'installer et de développer chez eux.

Lorsqu'on essaye de dépasser les bonnes intentions exprimées par les textes officiels définissant l'évaluation en Sciences pour voir les habitudes prises et la pratique réelle des concepteurs des différentes épreuves, on découvre que dans l'ensemble la pathologie de l'évaluation en Sciences dépasse la subjectivité longtemps dénoncée des jugements portés par les examinateurs pour se retrouver dans son manque chronique de validité par rapport aux contenus et aux objectifs déclarés. La validité étant par définition le degré auquel les épreuves mesurent ce qu'elles sont censées mesurer.

Dans de nombreux pays, l'enseignement est centré sur la structure des différentes matières qui constituent les disciplines.

On part le plus souvent du savoir universitaire déjà constitué pour confectionner une liste de thèmes, puis de contenus à enseigner qui occupent une place importante dans le système éducatif sous la forme de programmes dont l'élaboration peut ou non être confiée à une commission.

A partir de ce moment l'enseignement et l'évaluation de la discipline doivent impérativement se faire dans ce cadre, pour ne pas dire ce carcan pour respecter les directives officielles.

Dans le cas de l'évaluation sommative en Sciences, lorsqu'on effectue une catégorisation par thème des sujets proposés sur une assez longue période et qu'on la confronte au programme prescrit, on constate le plus souvent que dans la pratique tous les chapitres du programme ne sont pas logés à la même enseigne.

En effet dans la fréquentation des différents chapitres, on découvre l'existence d'une sorte d'échelle implicite de valeur qui fait qu'il y a des chapitres très fréquentés, d'autres qui le sont de temps à autre et une dernière catégorie qui est complètement abandonnée.

Parfois c'est à l'intérieur d'un même chapitre qu'il y a une nette discrimination entre les différentes notions et concepts.

S'agissant des objectifs testés, la situation est-elle aussi loin de la validité souhaitée?

Un dépouillement systématique des objectifs à partir des questions habituellement posées dans les épreuves de Sciences révèle que contrairement au désir affirmé et écrit de faire évoluer l'enseignement scientifique en direction de la démarche expérimentale et des attitudes scientifiques, ce sont les questions qui font appel à la restitution de connaissances qui l'emportent malgré les apparences trompeuses des différents libellés dont usent les concepteurs dans leur questionnement (Analyser, interpréter, déduire).

En réalité, ce qui est demandé aux élèves se résume à :

- des connaissances particulières,
- la possession d'une (ou de) stratégie(s) type(s) de résolution du problème,
- et exceptionnellement des capacités d'analyse de la situation proposée (JOHSUA, 1983).

Si nous prenons l'exemple de l'évaluation de la capacité des élèves à formuler des hypothèses, on constate que l'existence de lois connues dont la démonstration a nécessité des hypothèses connues des élèves parce qu'enseignées en classe ne laisse pas beaucoup de place à des questions testant la capacité de ces mêmes élèves à formuler des hypothèses originales.

L'expérience a montré que lorsqu'on essaie de poser de telles questions on tombe à coup sûr dans des questions de cours partielles et fragmentaires, qui n'insistent que sur certains points de raisonnement.

En effet, devant ces questions les élèves vont rechercher dans l'ensemble des explications qu'ils connaissent celles qui sont conformes puisqu'on leur demande toujours de sortir la bonne hypothèse et jamais d'exclure une hypothèse.

Finalement il s'agit seulement d'automatismes bâtis uniquement pour le contrôle des connaissances dans le cadre d'un système clos et sans autre pertinence que la réussite à l'examen (JOHSUA, 1983).

Quant aux objectifs affectifs, ils ne sont pas du tout évalués (il faut reconnaître qu'ils sont difficilement évaluables dans le cadre d'une épreuve papier crayon).

L'imagination dont font preuve les concepteurs de sujet ne fait que cacher mal la sclérose dans laquelle finissent par baigner la plupart des réformes.

Au bout d'un certain temps, l'examen tourne autour de questions types qui feront l'objet d'un bachotage tel que les demandes d'analyse, d'interprétation ou même d'explication feront appel à une restitution de connaissances.

Cette situation s'explique par le fait que tous les exercices proposés le sont dans le cadre d'un programme ; ce qui empêche de mettre l'élève en face d'un véritable problème, c'est-à-dire une situation tant soit peu nouvelle face à laquelle la réponse à donner doit être construite et pas du tout automatique.

Il y a également le fait que la formidable pression sociale qui s'exerce sur tout examen pousse inexorablement l'école vers une définition extensive de ce qui sera demandé aux élèves à l'examen.

D'ailleurs dans les pays où l'on a, à un moment donné, rêvé de pouvoir faire acquérir aux élèves, à propos des thèmes au programme, certains concepts de base et démarches à partir d'un ou deux exemples, pour ensuite lors de l'évaluation les amener à faire la preuve de leur apprentissage à travers une transposition dans un autre domaine non étudié en classe, on a vite déchanté. On a plutôt assisté à un phénomène insidieux d'inflation des connaissances dans les chapitres privilégiés par l'examen, puisque chaque fois qu'un exemple nouveau apparaissait aux examens, il faisait l'année suivante l'objet d'un apprentissage systématique en classe sur simple initiative des professeurs.

Finalement accablés par une grande quantité d'observations, de disciplines ou de définitions, la plupart des élèves ne trouvent pas d'autres solutions que d'apprendre par cœur autant qu'ils peuvent. La succession observation-mémorisation-contrôle-oubli devient pratique courante plutôt qu'observation-utilisation des concepts-interprétation-mise en relation avec des concepts plus vastes-résolution de problèmes (NOVAK, 1970).

Dans la plupart des cas, le programme de l'examen finit par remplacer celui réellement prescrit et l'ensemble du système enseignement-évaluation de tourner dans un cercle vicieux : on ne peut changer le programme prescrit parce qu'il prépare à l'examen, on ne peut changer l'examen parce qu'il sert à évaluer le programme !

Conclusion

Malgré les nombreux progrès de la docimologie, la situation de l'évaluation sommative en Sciences demeure encore un champ rempli d'incohérences et d'imperfections dont les conséquences sont énormes sur l'ensemble du curriculum.

Les programmes officiels ne sont plus respectés et sont calqués sur l'examen.

- Les manuels scolaires sont devenus très complexes et de niveau très élevé.
- Les professeurs enseignent de manière dogmatique et livresque pour pouvoir terminer les programmes, pire ils suppriment les travaux pratiques et se soucient très peu

de l'application des Sciences au monde qui entoure les élèves.

- Les élèves s'adonnent au bachotage systématique et réfléchissent peu. Même devant les situations nouvelles ils essaient coûte que coûte de placer leurs connaissances stéréotypées.

Face à cette situation, la tendance est de plus en plus à l'utilisation de nouveaux types d'épreuves et à la redéfinition des programmes.

Dans ce cadre, il est bon de se rappeler que quels que soient la philosophie et le contenu d'un programme d'enseignement, son efficacité réelle est en grande partie déterminée au bout de quelques années par les modalités de l'examen auquel les professeurs doivent préparer leurs élèves (GUINIER, 1980).

Dès lors les perspectives dans lesquelles il faut se situer sont les suivantes :

- Une redéfinition moins laconique et moins précipitée des programmes. Il est nécessaire pour cela de créer une réelle dynamique entre les contenus sélectionnés pour leur pertinence et les objectifs définis en termes de capacités réelles, c'est-à-dire de compétences spécifiques.
- La formation systématique des professeurs à la conception d'épreuves d'évaluation à plusieurs niveaux :
 - D'abord au niveau des objectifs de l'enseignement scientifique, il n'est pas normal que lorsqu'on demande à des enseignants d'indiquer les objectifs qu'ils poursuivent, la conversation tourne court (DE LANDSHEERE, 1975). Ils doivent pouvoir les expri-

mer clairement pour pouvoir savoir s'ils sont réellement évalués par les différentes épreuves.

- Ensuite, une formation à la conception et à la production d'exercices par un travail de déconstruction et de construction, mais aussi par l'analyse de copies d'élèves en vue de critiquer les exercices proposés dans leur validité et leur fidélité.

- Essayer de trouver d'autres formes d'organisation de l'évaluation qui permettent de minorer les contraintes sociales et institutionnelles qui ont tendance à enfermer les examens dans des questions types et des algorithmes.

Vaste programme certes, mais la réussite, nous semble-t-il, est à ce prix.

Bibliographie

BLOOM et al, (1969) - Taxinomie des objectifs pédagogiques. Lavallée trad. Montréal: Éducation nouvelle.

DE KETELE J.M. - ROEGIERS X. (1993) - Méthodologie du recueil d'informations. Bruxelles, De Boeck, p.74.

DE LANDSHEERE V. et G. (1975) . Définir les objectifs de l'éducation. Liège, Thone, p.15.

EBEL RL. (1965) - Measuring Educational Achievement (Englewood Cliffs N.J. Prentice Hall, Inc.), p.450.

FURST E.J. (1964) - Constructing Evaluation Instruments, New York David Mc Kay Co, Inc.

GUEYE B (1989) - L'épreuve écrite de Biologie au baccalauréat fait-elle appel au raisonnement en Sciences expérimentales ? ASTER N°8, pp.119-133.

GUINIER A. (1980) - À propos du contrôle des connaissances in Bulletin de l'Union des physiciens n° 627. pp.29-72.

HOTYAT F. (1962) - Les examens, les moyens d'évaluation dans l'enseignement. Paris, Bourrelrier.

JOHNSON S. (1983) - Contrôle des connaissances en fin de second cycle (baccalauréat) et nouveaux programmes de physique. Revue française de pédagogie n° 64, pp. 55-73.

MUNRO R.G. (1975) - In the structure of Sciences Education. Longman, Ed. P.L Gardner.

NOVAK J.D. (1970) . The improvement of biology teaching. New York, Bobbs Merrill and Company.

SCALLON G. (1988) - L'évaluation formative des apprentissages. Québec, Presses de L'Université LAVAL.

SCRIVEN M.S. (1967) . The methodology of evaluation (AERA Monograph series on curriculum evaluation, book 1). Chicago, Rand Mc Nally and Co.

STUFFLEBEAM D. et al (1980) - L'évaluation en éducation et la prise de décision. Ottawa, Editions NHP, p. 48.

TYLER W. (1950) - Basic principles of curriculum, and instruction, Syllabus for Education 360. Chicago, University of Chicago Press, p.69