

P10652B

290

Français
dans
le Monde n° 1000

vers
l'an
2000



Octobre-Novembre 1973

12 F

l'hypothèse favorable; ou, autrement, de déviations, de renoncements, voire de duperie.

L'évaluation n'aura guère de sens non plus si, par exemple, elle porte sur des opérations, des processus, des techniques, voire des domaines que l'enseignement n'aurait pas couverts. Combien de fois n'a-t-on pas vu un enseignement encyclopédique, impositif, centré sur la transmission de connaissances toutes faites, se terminer par un examen fait de questions dites d'intelligence, c'est-à-dire exigeant des applications, des transferts, des extrapolations auxquels l'élève n'a pas été entraîné. Puisqu'on interroge, dans ce cas, sur ce que l'on n'a pas enseigné, on suppose donc que la capacité de répondre provient d'aptitudes innées, d'une maturation et d'un développement spontanés. Mais qu'évalue-t-on alors ?

Enfin, que signifie une évaluation de connaissances pointillistes, comme elle se fait encore trop souvent, alors que l'on prétend aider l'étudiant à développer sa personnalité, sa créativité, sa tolérance, sa méthode de travail, son aptitude à travailler en groupe... ?

Traditionnellement, on s'est surtout attaché au produit (« qu'a appris l'élève ? »). En cas de succès, on croyait volontiers à la bonne qualité du maître et de son enseignement; l'échec inclinait plutôt à chercher les torts du côté de l'étudiant...

Les enseignants, gagnés dès maintenant à l'idée de l'évaluation permanente des élèves, vont devoir aussi s'habituer à l'évaluation permanente d'eux-mêmes, ainsi que des méthodes et des moyens qu'ils mettent en œuvre.

L'évaluation des maîtres se fera principalement par des systèmes d'analyse d'interactions, domaine qui, depuis quelques années, fait l'objet d'un gigantesque effort de recherche.

Qu'il s'agisse d'analyser les fonctions didactiques remplies par les interventions verbales des professeurs, la signification affective des comportements, verbaux ou non, ou encore une combinaison de ces aspects, la méthode reste fondamentalement la même. On prélève des échantillons d'enseignement (par magnétophone ou magnétoscope), lesquels peuvent être relativement courts tout en restant représentatifs de l'activité du maître considéré (quelques minutes suffisent, la plupart du temps). On analyse ensuite selon un système catégorique de complexité variable. N. Flanders, par exemple, a mis au point une démarche astucieuse qui, à partir de neuf catégories (accepte les sentiments de l'élève; encourage; accepte les idées des élèves; fait un cours *ex cathedra*; etc.), permet notamment d'évaluer le degré de centrage de l'enseigne-

ment sur la matière ou sur l'élève. Il n'est pas exclu que les maîtres disposent un jour dans leur classe d'appareils les informant du taux de participation verbale des élèves, de la durée moyenne des interventions de l'enseignant et des enseignés, de l'indice de centrage sur la matière ou sur les étudiants, etc. Des essais, dont on aurait tort de sous-estimer les apports, ont déjà été réalisés.

Nous avons proposé un système d'analyse rendant compte du temps consacré à l'organisation des situations d'enseignement, à l'imposition de connaissances, de méthodes, au développement des apports des élèves, à l'évaluation des réponses, aux interventions de nature affective, etc.². En France, les recherches de G. Ferry et de Prostic s'inscrivent dans cette direction. Enfin, dans le même ordre d'idées, la technique de micro-enseignement semble devoir jouer, à l'avenir, un rôle considérable, tant en ce qui concerne l'évaluation que le perfectionnement des maîtres de tous les niveaux.

Considérons maintenant trois aspects de l'évaluation des produits :

- L'évaluation du progrès de l'élève par rapport aux autres (*évaluation normative*). Les résultats sont, dans ce cas, comparés à ceux d'une population : la classe, ou l'ensemble des élèves de même niveau dans le pays ou dans la région, etc. Ainsi, l'élève se situe par rapport à la norme de son école, de sa région ou de son pays.

- L'évaluation du progrès de l'élève dans son propre apprentissage (*évaluation formative*). Les résultats sont comparés aux objectifs à atteindre. La maîtrise totale d'un apprentissage étant préalablement définie et cernée par des critères de réussite opérationnellement précisés, on situe l'étudiant dans son ascension vers l'objectif. L'apprentissage à effectuer constitue, en principe, une « unité de contenu ». Autrement dit, et comme nous allons le voir par la suite, les résultats sont évalués par rapport à une « population de questions » *représentatives* d'une unité de contenu.

- L'évaluation (comparée ou non) des programmes (*évaluation sommative*). Au niveau de l'élève, il s'agit du bilan des apprentissages réalisés sur plusieurs unités de contenu, ou plusieurs matières. Au niveau du *programme*, le bilan sera dressé en mesurant le rendement de l'enseignement fait à une population donnée d'élèves, sur une ou plusieurs unités de contenu.

2. Nous nous permettons de renvoyer à notre ouvrage : *Comment les maîtres enseignent*, Bruxelles, ministère de l'Éducation nationale, 1971, où l'on trouvera la description de notre système et une brève présentation des principaux systèmes américains.

Avant d'envisager l'évaluation formative, nous voudrions réagir contre le rejet abusif de toute forme d'examen où l'individu est jugé par rapport aux autres. Ce rejet procède ou bien d'une nouvelle forme de romantisme pédagogique (beaucoup plus répandue qu'on ne pourrait le croire), ou bien d'une volonté d'anarchie.

Pour orienter sa vie sociale, l'individu a besoin de savoir où ses diverses capacités le situent par rapport aux autres. Ce serait une duperie grave de laisser croire qu'une victoire sur soi-même ou sur un petit groupe de proches a valeur d'excellence universelle. Pour le bien de l'individu comme pour le bien commun, la société doit placer chacun là où il peut le mieux servir, au sens noble de cette expression, et donc le mieux se réaliser.

On l'aura compris, nous ne prôtons en rien le retour à l'examinite teintée de sadisme que pouvait se permettre une société viable avec l'exploitation réelle de quelque cinq pour cent de son potentiel intellectuel...

Méthodes et techniques

Qu'il s'agisse d'enseignement ou d'évaluation, un problème lancinant se pose à propos du français : « Est-il possible d'élaborer des programmes de langue maternelle séquentiels et cumulatifs, depuis le jardin d'enfants jusqu'à l'université? Un accord est-il possible sur un corpus de connaissances et un ensemble de techniques, et sur les étapes³ à y distinguer? » Moore et Kennedy remarquent : « Si nous ne pouvons pas répondre à cette question, nous devons nous résoudre à un avenir malheureux où l'actuel désordre des programmes persistera (...)⁴. »

Sans verser dans le travers d'une programmation qui se ferait carcan, sans vouloir imposer à chaque élève et à chaque maître une voie unique, inconciliable avec la nécessité impérative d'ajuster la relation élève-maître-matière selon les contingences du moment, il est néanmoins possible de définir des unités d'apprentissage à ordonner selon les circonstances. Nous ne croyons pas, en effet, qu'une analyse hiérarchique fine, complète et unique soit réalisable pour l'apprentissage de la langue. Par contre, il existe des grands carrefours, et les voies qui les séparent peuvent être jalonnées d'unités ordonnées *comme si* elles se répartissaient de façon critique,

3. National Council of Teachers, États-Unis. *Basic Issue Conference*, 1959. Cité par Moore, dans *Handbook*.

4. *Ibidem*.

5. Paris, E. Belin, 1939, 1^{re} partie.

6. Association Internationale pour l'Évaluation du Rendement Scolaire.

l'ordre pouvant d'ailleurs être à tout instant remis en question.

On travaille déjà dans ce sens depuis longtemps. Parmi bien d'autres exemples, citons la progression proposée par P. Clarac, B. Cognet et M. Janet dans leur manuel *Apprendre à écrire*⁵. L'étude de la description y comporte une succession de chapitres qui constituent aisément les départs d'unités d'apprentissage : Lumières vives - Lumières pâles - La nuit - La couleur - Les formes rectilignes - Les formes courbes, etc.

Assurément, cette démarche n'est pas la seule possible, mais elle fournit au professeur comme aux élèves des éléments d'ordre qui, exploités avec tout l'apport de la science et de l'art pédagogique d'aujourd'hui, semblent la voie la plus sûre vers un apprentissage fécond.

La notion d'unité d'apprentissage

Une unité d'apprentissage est toujours définie par l'objectif qui la commande. L'unité doit être conçue de façon à permettre une évaluation à court terme, la seule qui informe sans ambiguïté un maître déterminé de l'efficacité de son enseignement.

Cette évaluation à court terme ne suffit naturellement pas car les effets importants de l'éducation sont ceux qui durent, voire bonifient à l'école de la vie. Mais comment attribuer à l'influence d'un seul maître ce que l'on observe plusieurs années après son enseignement? En outre, les objectifs auxquels on accorde le plus d'importance ne peuvent être atteints que lentement. On n'apprend pas, en quelques mois, à s'exprimer clairement par la parole. Ici aussi, l'apport d'un maître isolé est pratiquement impossible à mesurer.

En fait, tel qu'il a été posé jusqu'à présent, le problème de l'évaluation des effets de l'enseignement à moyen et à long terme est sans solution. Une réponse ne sera possible qu'à la condition de faire systématiquement la distinction entre la responsabilité individuelle de l'enseignant, au niveau de l'unité d'apprentissage, et la *responsabilité collective* de tous les enseignants d'une même branche, à l'échelle du curriculum.

Des études de rendement et d'attitudes, dont les recherches comme celles de l'I.E.A.⁶ ont précisé la méthodologie, devraient être organisées à l'avenir, à intervalles réguliers, pour devenir finalement annuelles. Ces études de rendement seraient effectuées sur des échantillons d'élèves situés aux points clés de la scolarité (par exemple, 10, 15, 20 et 25 ans). On analyserait les

caractéristiques à la fois scolaires (type d'école, formation des maîtres, méthodes, etc.) et extrascolaires (milieu familial, *background* socioculturel). On pourrait ainsi déceler avec une rigueur rarement atteinte aujourd'hui les effets globaux du curriculum. Selon les résultats observés, des actions seraient entreprises auprès de l'ensemble des enseignants; des repères comportementaux précis leur seraient donnés pour modifier leur travail.

L'unité d'apprentissage, avons-nous dit, est définie par un objectif. Objectif défini *a priori*, *a posteriori*? La question est cruciale.

La période 1950-1970 a été marquée par l'apparition de taxonomies d'objectifs dont celle de B. Bloom et de ses collaborateurs est aujourd'hui la mieux connue⁷. La construction des tests et, plus généralement, des instruments d'évaluation trouvait ainsi un principe méthodologique d'un intérêt considérable. Partant d'un tableau à double entrée : objectifs - matières d'enseignement, le constructeur de tests découvre, à chaque intersection, la suggestion d'une question. Le tableau facilite en même temps la détermination de l'importance relative des différents objectifs et, pour chacun d'eux, des matières.

Exemple :

	Mémoire	Application
Énergie mécanique		
Gravitation		
Électricité statique		

Les trois principaux reproches formulés à l'égard de cette méthode sont les suivants :

- 1° Les objectifs spécifiques trouvés dans le tableau à double entrée ne sont pas nécessairement fidèles aux objectifs généraux de l'éducation. C'est le constructeur de tests qui croit, à tort ou à raison, qu'il en est ainsi;
- 2° Deuxième élément de subjectivité : la question posée jaillit de l'intuition du constructeur et non d'une règle sûre et automatique. Par exemple, pour l'ensemble des questions d'application portant sur la *connaissance de l'objet direct*, dix constructeurs de tests inventeront presque certainement dix *items* différents pour une même population;
- 3° Par cette méthode, la totalité des questions possibles n'est jamais formulée, voire définie. Il est donc impossible de démontrer que les questions posées effectivement constituent un échantillon représentatif de l'ensemble.

Ces trois éléments subjectifs ruinent en réalité le beau

rêve de rigueur scientifique des constructeurs de tests du troisième quart du XX^e siècle.

Comment s'orienter alors la recherche nouvelle? Le poids essentiel des efforts va, dans les prochaines années, porter sur la recherche de règles permettant de générer la totalité des questions d'un type donné.

Le premier exemple de réussite en ce sens nous est offert par le test de closure inventé par W. Taylor et dont nous avons récemment montré la validité pour la langue française⁹.

Traditionnellement, pour tester la compréhension d'un texte, l'évaluateur fait porter un certain nombre de questions sur ce qu'il croit nécessaire de vérifier (choix subjectif de la matière). De plus, il détermine, subjectivement aussi, un degré de difficulté des questions. En conséquence, une question formellement ou logiquement facile peut porter sur un texte difficile, tandis qu'une question difficile peut concerner un texte facile.

Avec ce genre d'épreuves, on ne sait jamais à partir de quel moment la question posée sort du donné actuel du texte pour mettre en jeu la mémoire de textes similaires, l'extrapolation, le jugement de valeur, etc.

Le test de closure semble éviter ces écueils. Il consiste à supprimer un mot sur cinq et à inviter l'élève à combler les lacunes; seul le mot originellement utilisé par l'auteur est admis comme correct. Dans une première forme du test, on supprime, par exemple, le 1^{er}, le 6^e, le 11^e mot, etc.; dans une deuxième forme, on supprime le 2^e, le 7^e, le 12^e mot, etc. Avec cinq formes différentes, *tous* les mots du texte, c'est-à-dire l'univers complet des questions, sont couverts.

7. B.S. BLOOM et al., *Taxonomie des objectifs pédagogiques. Domaine cognitif*, Montréal, Éditions nouvelles, 1969; *id.*, *Domaine affectif*, 1970.

Pour le domaine cognitif, il s'agit d'une hiérarchie d'objectifs dont le plus élémentaire est la mémorisation et la restitution de faits isolés, et dont le plus élevé est la capacité de porter des jugements en fonction de critères objectifs ou subjectifs.

Exemple (Davis) : Matière = compréhension de la lecture.
Objectifs

Niveau I	Questions portant sur le sens d'un mot.
Niveau II	Questions portant sur l'organisation d'un passage.
Niveau III	Questions portant sur des informations contenues dans le texte.
Niveau IV	Questions portant sur des informations devant être inférées à partir du texte.
Niveau V	Questions portant sur le point de vue de l'auteur.

9. G. DE LANDSHEERE : *Le test de closure. Mesure de la compréhension de la lisibilité*, Paris, Nathan; Bruxelles, Labor, 1973.

Cette épreuve, d'apparence naïve ou étroite, permet, en réalité, une bonne mesure de la compréhension et de la lisibilité. Elle discrimine bien les niveaux scolaires entre eux.

A côté de cet exemple de simplicité parfaite, comment la recherche s'orienté-t-elle ?

J.R. Bormuth écrit¹⁰ : « Une définition opérationnelle d'une classe de questions de tests de connaissances est une série de directives indiquant au rédacteur de questions comment il doit réarranger des segments d'enseignement pour obtenir des questions du type désiré. » A aucun moment, ces directives ne doivent faire appel à l'introspection du rédacteur.

Le constructeur de questions va donc, en quelque sorte, « coller » à la matière. C'est de celle-ci et d'elle seule, que dériveront les questions posées; en fait, chacune sera créée en appliquant à un fragment de cette matière des règles systématiques. Soit, par exemple, le fragment de matière : « C. Colomb découvre l'Amérique en 1492. » Une même règle appliquée successivement aux différents constituants de la phrase permet de formuler l'univers de questions suivant :

- Qui découvre l'Amérique en 1492 ?
- Que découvre C. Colomb en 1492 ?
- Quand C. Colomb découvre-t-il l'Amérique ?
- Que fit C. Colomb en 1492 ?

L'exemple peut paraître simpliste. La technique proposée par J. Bormuth ouvre néanmoins des perspectives considérables. Par le biais du langage, on peut espérer dresser un inventaire — à la limite exhaustif — des contenus que ce langage véhicule.

La linguistique assume ici un rôle d'outil de première importance et apporte des hypothèses de travail d'une grande fécondité.

Revenons un instant, à titre d'illustration, à l'évaluation de la compréhension de la lecture. L'échelle de niveaux sera remplacée par une structuration à base linguistique :

- Compréhension du mot;
- Compréhension de la structure de la phrase;
- Compréhension des relations anaphoriques;
- Compréhension des relations inter-phrases et inter-paragraphe;
- etc.

A chacune de ces aires correspondront des règles de génération de questions. Pour mesurer la compréhension d'un texte, par exemple : *La cigale et la fourmi*, on posera à l'élève un échantillon représentatif de la population de questions pouvant être générées par ces règles.

— La compréhension lexicale pourrait être testée en relevant un échantillon des mots du texte qui n'appartiennent pas à un vocabulaire de base donné et en construisant pour chacun un même type de question (fondée, par exemple, sur un ensemble de mots appartenant au même champ sémantique).

- La bise, c'est : — la pluie ?
— la neige ?
— le vent ?
— le brouillard ?

— Un certain nombre de phrases seront tirées au sort, et on construira, pour chacune, des questions testant la compréhension des structures.

- Qui se trouve fort dépourvue ?
- Quand la cigale se trouva-t-elle fort dépourvue ?
Etc.
- « Son *moindre* défaut », c'est « son défaut le plus petit » ?
ou « son défaut le plus grand » ?

— L'analyse du réseau anaphorique qui traverse le texte permet de constituer un échantillon de questions testant la compréhension de ces relations.

- Qui dit : « Je vous paierai » ?
- Qui est « cette emprunteuse » ?
Etc.

— Diverses techniques peuvent être mises au point pour évaluer la compréhension inter-phrases. La succession temporelle ou les relations causales pourraient donner lieu à des questions où des phrases ou des paragraphes présentés dans le désordre doivent être réarrangés.

— De nombreux autres aspects pourront naturellement être explorés selon le même principe.

Au service de cette nouvelle technique (qui commence à être connue en Europe, sous son nom américain de « tests à référence critérielle » (*criterion referenced tests*), des outils de contrôle statistique, nouveaux, eux aussi, sont ou vont être forgés : contrôle de l'homogénéité des questions issues d'une même règle, détection des hiérarchies existant entre les questions issues de deux ou plusieurs règles.

10. J.R. BORMUTH : *On the Theory of Achievement Test Items*, University of Chicago Press, 1970.

Un corollaire important du principe de génération automatique des questions est, en théorie, l'absolue répliquabilité du test : deux formes de l'épreuve, fondées sur des échantillons de questions issues des mêmes règles, sont représentatives de la même population de questions, et donc exactement comparables. Ce qui fait des tests à référence critérielle un outil particulièrement précieux pour l'évaluation continue.

On aurait tort de croire que de telles considérations théoriques ne peuvent pas servir la pratique quotidienne. Bien mises au point, les règles de génération de questions se révèlent, au contraire, d'une grande simplicité et d'une efficacité élevée.

Que deviennent les objectifs, dans cette perspective? Ils restent naturellement l'aspect essentiel. Mais la démarche habituellement suivie pour la construction des examens est partiellement inversée. Les objectifs généraux, puis intermédiaires, étant définis, on ne tente plus de rédiger directement un exercice ou une question qui y corresponde. On passe d'abord en revue les types de questions applicables et ce n'est qu'après avoir acquis la conviction qu'une correspondance objectif type d'évaluation existe bien que l'on arrive aux questions particulières, dont un ou plusieurs paramètres sont fixés.

Il n'est probablement pas exagéré de penser que la technique de génération automatique de questions marque un tournant décisif dans l'histoire de l'enseignement et de l'évaluation.

L'évaluation sommative

En se situant dans une unité d'apprentissage, l'évaluation formative prend fondamentalement un caractère diagnostique.

Mais, en additionnant les évaluations portant sur un ensemble d'unités constituant un *programme*, on dresse en même temps un bilan des apprentissages. On est ainsi passé en système sommatif.

Comme on l'a déjà indiqué, l'évaluation sommative peut concerner soit un élève particulier, soit un programme d'enseignement, jugé d'après les performances d'une population.

J. Bormuth, de nouveau lui, a peut-être le mieux cerné le problème théorique crucial de la comparaison des *curricula*.

Jusqu'à présent, en effet, on a tenté des évaluations comparatives de méthodes d'enseignement ou de programmes sans veiller strictement à la comparabilité des éléments pris pour critères.

Parmi les conditions nécessaires à la comparaison, nous retiendrons spécialement :

1. La relation fixe d'un type de question donné avec le programme doit être établie et les règles de génération doivent être définies;
2. La comparaison doit se faire entre questions de même type;
3. Pour chaque type, le niveau de représentativité des questions posées doit être établi;
4. Si certains programmes permettent des types de questions et d'autres pas, les questions tombant dans cette catégorie doivent être traitées séparément.

Vers une individualisation de l'évaluation

Les progrès technologiques garantissent, dès maintenant, la possibilité d'individualiser l'enseignement. Mais les tests de connaissances actuels couvrent un spectre si large et sont étalonnés sur des populations tellement vastes que les normes ne sont guère utilisables dans des situations d'apprentissage restreintes.

Aux tests conçus pour une région ou pour une nation, il faut substituer des tests spécifiques à une classe ou à un groupe de travail. Ce problème est, dès maintenant, résolu dans son principe, et des solutions sont déjà opérationnelles : ce sont les *banques d'objectifs et de questions*.

Dans une telle banque, les informations sont classées selon les objectifs, les matières et les types de questions. Pour chaque question, on rassemble peu à peu les données de rendement pour différentes populations et les résultats d'analyse clinique des réponses (les indications psychométriques usuelles sont aussi stockées¹¹).

Dans l'hypothèse où une école est reliée à une banque de questions par un *terminal* d'ordinateur, tout enseignant peut (au moment où le système est complètement développé) demander à tout instant un test de *x* questions de tels types, explorant tels objectifs dans telle matière, pour tel niveau. Dans des conditions de gestion automatique, l'enseignant voit apparaître, quelques instants après, le test spécialement conçu pour ses élèves. En système moins automatisé, le temps nécessaire se limite à l'aller et retour du courrier postal.

11. Une telle banque, commencée il y a deux ans par D. Leclercq, assistant à notre Laboratoire, est maintenant en service à l'École de la Force aérienne belge, à Saffraenberg.

Et les pronostics d'apprentissage?

Toute évaluation s'ouvre sur un pronostic. Car l'enseignement exige un ajustement incessant, en fonction de la facilité ou de la difficulté de l'apprentissage qui se prépare.

Dans les quelques considérations qui suivent, nous voudrions cependant nous limiter aux pronostics à moyen terme, nécessaires aux points clés de l'enseignement.

En raison du flou des programmes actuels, de leur absence de linéarité, beaucoup de pronostics sont pratiquement impossibles. On ne peut prédire que les choses opérationnellement définies. Nous sommes ainsi ramenés à la définition des objectifs et des chemins critiques.

L'exemple suivant montre quel degré de précision on atteint déjà dans la prédiction de certains apprentissages.

On connaît les travaux d'Inizan¹² sur la prédiction de la réussite ou de l'échec en lecture à partir d'une batterie de tests administrée aux enfants avant leur entrée en première année primaire. En les développant quelque peu, en affinant les analyses de régression et, surtout, en les validant par des observations longitudinales, une de nos collaboratrices a réuni des données suffisamment précises pour permettre une automatisation des traitements¹³. Aujourd'hui, l'ordinateur trace le profil d'apprentissage un an d'avance et précise à l'intérieur de quelle marge de succès l'élève doit normalement progresser. Dès que le profil observé descend sous le profil prédit, on peut pratiquement être sûr qu'un accident s'est produit. Il importe naturellement de le rechercher et d'apporter les remèdes nécessaires.

Nous croyons que de semblables procédures vont se multiplier dans les décennies prochaines.

Ce regard prospectif sur l'évaluation serait trop incomplet si nous ne disions pas un mot de l'évaluation de la difficulté des textes, si importante pour l'apprentissage de la langue.

12. INIZAN : *Le temps d'apprendre à lire*, Paris, Bourrellet, 1964, 2^e éd.

13. E. LECLERQ-BOXUS : *Prédiction de l'apprentissage de la lecture en première année primaire. La méthode PREDIC*, Laboratoire de Pédagogie expérimentale de l'Université de Liège.

14. Elles sont dues à G. HENRY, *Une technique de mesure de la lisibilité spécifique à la langue française. Trois types de formules adaptées à trois niveaux d'enseignement*, Université de Liège, à paraître.

Les formules de premier type sont lourdes et uniquement destinées à la recherche; celles du second type sont automatisables et exigent, pour l'usage courant, le recours à l'ordinateur; celles du troisième type sont utilisables par chacun, sans aide technologique.

Depuis cette année, il existe enfin des formules de lisibilité spécifiques à la langue française¹⁴. Elles sont, elles aussi, nées de la linguistique contemporaine. Une nouvelle technologie de la lecture est en train de se développer. On peut, dès maintenant, être sûr que, d'ici quelques années, chaque lecteur pourra être servi dans les bibliothèques, non seulement en fonction de ses intérêts, mais aussi de ses capacités lexiques et des objectifs poursuivis.

Le message verbal et aussi l'audio-visuel feront l'objet de démarches similaires. Nous les voyons se dessiner de plus en plus nettement.

En guise de conclusion

Avons-nous réussi à donner une image de l'évaluation du français en l'an 2000? Certes non. Un gros livre eût à peine suffi à l'entreprise.

Dans l'impossibilité d'approfondir le thème proposé, nous nous sommes arbitrairement limité à quelques aspects marquants.

A partir du moment où l'évaluation est conçue comme partie intégrante de l'enseignement, faire la prospective de l'une exige aussi la prospective de l'autre.

Nous sommes, dès maintenant, convaincu que, dans les prochaines années, l'évaluation atteindra un degré de précision et une rigueur qu'on ne soupçonne pas encore. Elle accompagnera la transformation profonde que la pratique scolaire s'apprête à subir. Cette transformation constituera un aboutissement (provisoire) de l'extraordinaire mouvement de recherche en éducation qui s'est développé ces vingt dernières années, mais n'avait encore trouvé ni sa synthèse, ni sa voie de pénétration effective dans notre pédagogie.

La recherche sur les programmes et les méthodes d'enseignement connaîtra de son côté une révision parfois déchirante, car bien des résultats tenus pour acquis vont se voir brutalement contestés.

J. Bormuth écrit : « Si l'expérimentateur ne peut pas vérifier si ses questions sont bien du type qu'il prétend et si d'autres chercheurs ne peuvent pas construire des questions qu'ils puissent certifier du même type que celles de leur collègue, ces chercheurs ne peuvent en rien prétendre qu'ils ont réfuté ou confirmé les résultats du premier. En pareil cas, l'évaluation originale est sans valeur. »

G. De Landsheere

