

CE 1886.

672789

ouvrage  
RS, Paris.  
Nancy II.  
bonne (Paris V).  
Liège.  
Genève.  
ie de Caen.  
de Paris V.  
Paris V.  
Caen.  
S, Paris.  
trique, Université de Paris V.  
Caen.

# TRAITÉ DES SCIENCES PÉDAGOGIQUES

PUBLIÉ SOUS LA DIRECTION DE  
MAURICE DEBESSE ET GASTON MIALARET

7

*fonction et formation  
des enseignants*

*par F. Aniel-Lebigre, J. Berbaum, M. Debesse  
G. De Landsheere, A.-M. Huberman, D. Lavenu  
L. Lefèvre, A. Léon, G. Mialaret, S. Mollo  
P. Pichot et J. Vial*



PRESSES UNIVERSITAIRES DE FRANCE  
108, BOULEVARD SAINT-GERMAIN, PARIS

UNIVERSITÉ DE LIÈGE  
Service de Psychologie de  
l'Enfant et de l'Adolescent  
BIBLIOTHÈQUE

459

## Préface

PAR MAURICE DEBESS

*La genèse de ce septième volume en explique à la fois le titre, paru, à la réflexion, que son lecteur.*

*Dans la conception d'un volume serait consacré à la F il était apparu que les problèmes de l'information culturelle qui s'inscrivent dans des spécialités méritaient un traitement en deux volumes se complétant. La préparation de l'un d'eux*

*Celle du présent volume traite tout d'abord : l'étude de la formation seule tout un traité, si l'on veut, qu'on retrouve aussi bien dans l'éducation scolaire et dans la formation de se limiter à la formation pédagogique où l'accent est mis sur l'éducation scolaire et présente où leur formation peut bénéficier d'une disposition par les mass media*

*Ce n'est pas sans regret que l'éducateur est plus et mieux informé. Il n'y a pas de véritable enseignement de l'éducateur, c'est-à-dire celle*

Dépôt légal. — 1<sup>re</sup> édition : 1<sup>er</sup> trimestre 1978  
© 1978, Presses Universitaires de France  
Tous droits de traduction, de reproduction et d'adaptation  
réservés pour tous pays

spective, 1970, 204-7.  
*maîtres enseignants. Analyse des inter-*  
*ducation nationale et de la Culture,*  
*édagogiques : moyen de diagnostic*  
*française de Pédagogie, janv.-févr.-*  
 Paris, PUF, 1964, 298 p.  
*établissement scolaire. Quatre enquêtes*  
 1, 1969, 222 p.  
 is of aggressive behavior in experi-  
*psychol.*, 1969, 10.  
 eur par la méthode des incidents  
 pt. 1956, XVIII, n° 75, p. 69-88.  
*eurs de mathématiques*, Paris, CNRS,  
 1959, 77 p.  
*iques*, Paris, PUF, 1964, 462 p.  
 mation des éducateurs, in *Sciences*  
 8.  
 Paris, PUF, 1974.  
 on des éducateurs, *Les Sciences de*  
 1973, p. 183-211.  
 coll. « Que sais-je ? ».  
 1970, coll. « L'École et la Société »  
 (iers).  
 , rééd. 1971, « Que sais-je ? ».  
 es professeurs de sciences, in *Les*  
 , p. 57-119.  
*ogique des maîtres*, Rouen, INRDP,  
 PUF, 1971, 126 p.  
*iques et pédagogiques de la formation*  
 graphié, Caen, 1973.  
 e, 1975, 135 p.

## L'évaluation des enseignants

PAR GILBERT DE LANDSHEERE

### Sommaire

INTRODUCTION	111
I - QU'EST-CE QU'UN BON ENSEIGNANT?	112
Bon professeur pour qui?	113
II - QUALITÉS NÉCESSAIRES DE L'ÉVALUATION	115
1 - La validité	115
Evaluation formative	115
Evaluation normative	118
2 - La fidélité	118
III - LES MÉTHODES ET LES TECHNIQUES	120
1 - L'étude des caractéristiques	120
2 - L'étude descriptive des processus d'enseignement	122
Les systèmes de signes	123
Les systèmes de catégories	126
3 - L'évaluation des produits	129
IV - QUI ÉVALUE?	135
1 - L'évaluation par les inspecteurs	135
2 - L'évaluation par les collègues	136
3 - L'évaluation par les élèves	137
4 - L'auto-évaluation	141
CONCLUSION	142
<i>Bibliographie</i>	143

## Introduction

Faut-il évaluer les enseignants — comme l'annonce le titre du présent chapitre — ou leur enseignement ? La question n'est pas oiseuse. *A priori*, on peut penser que seule importe l'efficacité avec laquelle la fonction est remplie, indépendamment des caractéristiques, des traits de personnalité de l'enseignant. J. McNeil et W. Popham (1973, p. 219) écrivent avec raison :

« Si l'on continue à poser que bon enseignant et bonne personne sont des termes équivalents, les problèmes relatifs à l'efficacité des maîtres resteront intraitables et les efforts déployés en la matière seront vraisemblablement improductifs. »

La confusion entre ce qui relève de la personne humaine (domaine du privé) et l'enseignement (fonction publique) trouve son explication essentielle dans le fait que le concept d'éducation est indissociable de l'idée de progrès cognitif, affectif, psychomoteur. Et nous répugnons à la pensée qu'un progrès de l'éducation puisse être attribuable à une personne perverse, désordonnée, instable...

De façon plus circonstancielle, on se résout souvent à apprécier la personne faute de pouvoir évaluer avec quelque sûreté les effets de son action.

Quatre raisons majeures incitent à évaluer l'enseignant ou l'enseignement :

1 / Avant d'engager un éducateur, de donner un caractère définitif à son emploi, de lui accorder une promotion, on désire s'assurer de sa qualification, de sa compétence (raison administrative).

2 / On ne peut améliorer systématiquement l'enseignement sans disposer de critères de progrès.

3 / Pour former les futurs maîtres, il est nécessaire de pouvoir distinguer avec quelque certitude le bon enseignement du mauvais.

4 / Enfin, les effets mesurés de l'enseignement constituent la variable dépendante dans de nombreuses recherches.

## I

## Qu'est-ce qu'un bon enseignant ?

Dans l'abstrait, le meilleur enseignant est celui qui suscite chez ses élèves la plus grande quantité d'apprentissages de la meilleure qualité.

Mais de quels apprentissages s'agit-il ? Ils peuvent être cognitifs, affectifs ou psychomoteurs, simples ou complexes, durables ou éphémères. On admet généralement qu'il existe des processus intellectuels plus nobles que d'autres. L'analyse, la synthèse, la résolution de problèmes, la créativité occupent, dans la hiérarchie des processus intellectuels, une place supérieure à la mémoire. On sait aussi l'importance des attitudes et des valeurs, et beaucoup d'entre elles ne s'installent et ne se modèlent ou ne se remodelent que lentement : conquête de l'indépendance, tolérance, sens social, curiosité d'esprit...

Un professeur n'ambitionne pas de provoquer seul tous ces apprentissages, mais il ne peut pas, non plus, limiter arbitrairement son action à un seul secteur (qui fut trop souvent celui des connaissances).

Certains apprentissages doivent, par ailleurs, se conjuguer : conquérir le savoir et éprouver un intérêt, un goût croissant pour le domaine que l'on a choisi d'étudier...

Fait capital, les maîtres (comme leurs élèves) ne sont jamais des êtres neutres, susceptibles d'être entièrement façonnés par des théories ou des pratiques pédagogiques ou autres. Ils sont doués d'une personnalité en partie invariante, et cette personnalité détermine à tel point le style de l'enseignement qu'à partir de nombreux travaux (Bush, 1954; Ryan, 1960-1972; Rosenshine, 1971b; Postlethwaite, 1974) on arrive à une typologie. Elle s'établit *grossomodo* de la façon suivante :

1° *Type X*. — S'attache au développement de la personnalité de l'étudiant, en accordant une grande importance à l'affectif, au social. Suit un programme souple, sans trop se soucier de la matière couverte. Méthode d'enseignement sans cérémonie, aussi individualisée que possible. Chaleureux, amical.

*Ant.* : distant, égocentrique, réservé.

2° *Type Y*. — Se soucie uniquement de l'intelligence de ses élèves. S'attache au sujet traité. Suit un programme détaillé, logique. Fixe

des normes élevées. Exerçant un pouvoir de contrôle. Distant, il ne noue avec ses élèves de relations professionnelles.

*Ant.* : inorganisé, négligent.

3° *Type Z*. — Stimule l'initiative de ses élèves, qui donnera à un petit nombre d'entre eux l'occasion d'exprimer leur génie créatif. Réactions envers les élèves et collègues personnels.

*Ant.* : terne, routinier.

X, Y et Z se rencontrent dans la pratique de façon de dominantes plus ou moins égales. X est le type de « bons » professeurs ; Y favorise les connaissances ; Z favorise les réactions (Popper, 1964). X, Y et Z négatifs ? Ce n'est même pas la question.

## BON PROFESSEUR POUR QUOI ?

On n'est pas bon professeur dans toutes les situations (y compris la situation de classe). On est exceptionnel de convenir que dans des situations de vastes auditoriums ou de petits séminaires de licence.

« La qualité de l'apprentissage est une donnée de l'enseignement donnée par un professeur *particulier* pour atteindre un but *particulier* » (Popper, 1964).

Il serait donc injuste de faire des jugements relatifs à des enseignements différents. Qu'y a-t-il encore de bon pour un professeur de propédeutique ou de philosophie et un professeur hautement qualifié au troisième cycle à quel point ?

La façon dont les enseignants agissent d'ailleurs plus de leur personnalité que de leurs pédagogies concrètes.

« Il n'existe pas, dans la pratique, de méthodes généralement acceptées vis-à-vis de l'évaluation des enseignants ».

des normes élevées. Exerce des contrôles de connaissances rigoureux. Distant, il ne noue avec les étudiants que des relations de type professionnel.

*Ant.* : inorganisé, négligent.

3° *Type Z*. — Stimulant, imaginaire. Essaie de provoquer l'étincelle qui donnera à un petit nombre d'élèves l'énergie, la liberté et l'art d'exprimer leur génie créateur. Ne se soucie pas étroitement de l'intelligence au sens traditionnel. Exerce un contrôle de rigueur variable. Réactions envers les élèves souvent dictées par des sentiments personnels.

*Ant.* : terne, routinier.

X, Y et Z se rencontrent rarement à l'état « pur » ; il s'agit plutôt de dominantes plus ou moins marquées. X, Y et Z peuvent tous être de « bons » professeurs. Y assure le mieux l'apprentissage de connaissances ; Z favorise plus la compréhension (Solomon *et al.*, 1964). X, Y et Z négatifs sont-ils nécessairement de mauvais professeurs ? Ce n'est même pas certain, au moins à certains égards.

#### BON PROFESSEUR POUR QUI ?

On n'est pas bon professeur dans l'absolu, mais en fonction de la situation (y compris la matière à enseigner) et des étudiants. Il est exceptionnel de convenir à tous. Tel professeur se sent stimulé par de vastes auditoriums alors que tel autre ne se révèle que dans des petits séminaires de licence.

« La qualité de l'apprentissage que l'on pressent dans une situation d'enseignement donnée est fonction d'une méthode *particulière* employée par un professeur *particulier* pour des étudiants *particuliers* poursuivant un but *particulier* » (Popham, 1967).

Il serait donc injuste et irréaliste de comparer les évaluations relatives à des enseignants travaillant dans des conditions très différentes. Qu'y a-t-il encore de commun, par exemple, entre un professeur de propédeutique qui s'adresse à plusieurs centaines d'étudiants et un professeur hautement spécialisé, ne faisant que des cours de troisième cycle à quelques chercheurs ?

La façon dont les étudiants perçoivent un professeur dépend d'ailleurs plus de leur système de valeurs que de caractéristiques pédagogiques concrètes :

« Il n'existe pas, dans notre société, un ensemble d'attentes universellement acceptées vis-à-vis du professeur. Il existe, au contraire,

un nombre inconnu d'ensembles différents, dépendant du fonds culturel et socio-économique des individus. Pour prédire comment un observateur jugera un professeur particulier, il faut d'abord découvrir quel rôle l'observateur attend de l'enseignant » (A. Cohen *et al.*, 1973).

Turner (1965) fait, dans ce contexte, une observation frappante (qui appelle certainement des vérifications) : dans des classes fréquentées par une majorité d'enfants d'ouvriers, la façon d'enseigner, plus que les caractéristiques personnelles et sociales des maîtres, permettrait de prédire le succès de l'enseignant. Si les enfants de classe moyenne dominant, les caractéristiques personnelles et sociales des maîtres seraient plus déterminantes que la façon d'enseigner.

De façon plus générale, il existe une relation maintes fois démontrée entre les caractéristiques de la personnalité d'un étudiant et l'avis qu'il porte sur un enseignant. En particulier, Morrison (1966) distingue les élèves selon leur degré de sensibilité aux renforcements extérieurs (extériorité) et remarque que les étudiants possédant ce trait à un degré élevé réagissent, beaucoup plus que les autres, aux changements de style d'enseignement. Bien que les résultats de la recherche manquent de cohérence, il semble aussi que les étudiants doués ne jugent pas de la même façon que leurs compagnons plus faibles. Spaight (1967) note qu'un même professeur est jugé froid, distant par 12 % d'étudiants brillants et par 94 % d'étudiants faibles. Enfin, on constate (rarement, il est vrai) certaines différences d'évaluation selon le sexe de l'étudiant. Les étudiantes semblent accorder plus d'importance que leurs compagnons à une bonne structuration de la matière (McKeachie, 1971).

Il n'existe donc pas de bon professeur dans l'absolu. Faut-il pour cela renoncer à toute évaluation ? Assurément non. D'abord parce que, avec une méthode particulière, le professeur particulier, s'adressant à des étudiants particuliers, pour poursuivre un but particulier... atteint ce but ou non, et il est important de le savoir. Ensuite, parce que toutes ces restrictions peuvent être prises en compte pour créer des situations expérimentales permettant de comparer des enseignants attachés, non pas à des tâches strictement pareilles, mais relativement équivalentes. Enfin, on constate que, malgré la mouvance des situations, certains enseignants atteignent généralement mieux l'objectif fixé que d'autres. Les inspecteurs expérimentés le savent depuis longtemps, mais il appartient à la recherche de préciser les choses...

## Qualités

## I | LA VALIDITÉ

Pourquoi évalue-t-on des buts que lui assigne le même, ou pour comparer ?

On retrouve ainsi la *évaluation formative* et l'*évaluation de validité*, la première n'étant pas un générateur de progrès. La sélection et certaines s'excluent pas mutuellement.

*Evaluation formative*

Le but de l'enseignant est de proposer des apprentissages propices à des apprentissages donc :

- 1° que l'enseignant sache les connaissances et les nécessaires et sache les
- 2° qu'il propose des objectifs dans la ligne des fins même et ses élèves accèdent
- 3° que les contenus proposés soient exacts ;
- 4° que les élèves apprennent

Avant de porter un jugement sur les aspects devraient être comparés est satisfaite, les autres inconvénients s'attacher qu'à l'une des de la mieux remplir.

1° Créer les conditions multiples incertitudes qu'elles sciences de l'éducation, il

dépendant du fonds  
 Pour prédire comment  
 culier, il faut d'abord  
 enseignant » (A. Cohen

observation frappante  
 : dans des classes fré-  
 rs, la façon d'enseigner,  
 t sociales des maîtres,  
 nant. Si les enfants de  
 personnelles et sociales  
 la façon d'enseigner.

ion maintes fois démon-  
 alité d'un étudiant et  
 culier, Morrison (1966)  
 ilité aux renforcements  
 étudiants possédant ce  
 plus que les autres, aux  
 que les résultats de la  
 e aussi que les étudiants  
 leurs compagnons plus  
 professeur est jugé froid,  
 94 % d'étudiants faibles.  
 taines différences d'éva-  
 antes semblent accorder  
 une bonne structuration

ns l'absolu. Faut-il pour  
 ent non. D'abord parce  
 sseur particulier, s'adres-  
 ivre un but particulier...  
 le savoir. Ensuite, parce  
 es en compte pour créer  
 comparer des enseignants  
 reilles, mais relativement  
 é la mouvance des situa-  
 alement mieux l'objectif  
 entés le savent depuis  
 herche de préciser les

## II

### Qualités nécessaires de l'évaluation

#### I | LA VALIDITÉ

Pourquoi évalue-t-on ? Pour savoir si un enseignant atteint les buts que lui assigne le programme scolaire ou qu'il s'est fixés lui-même, ou pour comparer sa performance à celles de ses collègues ?

On retrouve ainsi la distinction aujourd'hui classique entre l'*évaluation formative* et l'*évaluation normative*. Outre sa plus haute probabilité de validité, la première revêt normalement un caractère diagnostique générateur de progrès. La seconde présente surtout un intérêt pour la sélection et certaines promotions. Les deux types d'évaluation ne s'excluent pas mutuellement.

#### *Evaluation formative*

Le but de l'enseignant est de créer ou de préserver les conditions propices à des apprentissages souhaitables chez les élèves. Il importe donc :

- 1° que l'enseignant sache créer ces conditions, c'est-à-dire possède les connaissances et les habiletés psychologiques et pédagogiques nécessaires et sache les mettre en œuvre;
- 2° qu'il propose des objectifs inscrits directement ou indirectement dans la ligne des fins et des buts de l'éducation, et auxquels lui-même et ses élèves adhèrent;
- 3° que les contenus proposés aux élèves ou mis à leur disposition soient exacts;
- 4° que les élèves apprennent en se développant positivement.

Avant de porter un jugement général sur un enseignant, ces quatre aspects devraient être considérés. A la limite, si la quatrième condition est satisfaite, les autres importent peu, mais il est parfois indiqué de ne s'attacher qu'à l'une des trois autres pour s'efforcer systématiquement de la mieux remplir.

1° *Créer les conditions propices aux apprentissages.* — Au-delà des multiples incertitudes qui continuent à grever la psychologie et les sciences de l'éducation, il existe un acquis dans le domaine de l'hygiène



physique et mentale, du développement de la personnalité, de l'apprentissage, de la psychologie des groupes, de la méthodologie de l'enseignement que tout éducateur devrait dominer. Certaines exigences cataloguées comme administratives (ponctualité, tenue régulière des dossiers scolaires) appartiennent aussi à la catégorie des conditions.

2° *Les objectifs poursuivis.* — Faute de conduire un train fou, ou de croire à la génération spontanée des apprentissages, un enseignant doit toujours pouvoir expliquer les raisons de son action, qu'il poursuive des buts à large spectre ou des objectifs très spécifiques. La méthodologie de la définition des objectifs en éducation constitue d'ailleurs un des secteurs où la pédagogie a le plus progressé au cours de ces dernières années (V. et G. De Landsheere, 1975).

Du sommet à la base d'une hiérarchie pédagogique, les objectifs tendent à se diversifier. Même en présence d'une idéologie unique, clairement définie au départ, et même si les enseignants adhèrent profondément aux buts assignés, ils n'en procèdent pas moins à une interprétation personnelle, consciente ou inconsciente.

Selon qu'il appartient au type X, Y ou Z, un enseignant privilégie certains objectifs cognitifs ou affectifs. Evaluer un enseignant dans l'absolu, avec un instrument à spectre étroit, ne revêt donc guère de sens. Ou bien l'évaluateur précise qu'il ne s'intéresse qu'à certains comportements et situe son évaluation dans ce cadre, ou bien il change d'instruments selon le type d'enseignant auquel il a affaire et, plus spécialement encore, selon les objectifs précis que le professeur déclare poursuivre. On peut naturellement ne pas être d'accord sur les objectifs mêmes; dans ce cas, il ne s'agit plus d'évaluation d'efficacité à proprement parler, mais plutôt d'adéquation entre les fins, les buts et les objectifs de l'éducation.

3° *L'exactitude et l'adéquation des contenus.* — On peut enseigner avec talent et efficacité des choses inexactes, humainement mutilantes ou parfaitement inutiles. On tient généralement pour évident que la justesse des contenus est toujours évaluée; ce n'est pas sûr. En outre, à partir d'un certain niveau de connaissances ou de complexité psychologique, la qualification des évaluateurs devient de plus en plus rare. Dans l'enseignement supérieur, notamment, seule une appréciation par les pairs reste possible.

Dès que l'on aborde le domaine des valeurs et des attitudes, les croyances prennent le pas sur la science, et la validité devient indémontrable universellement. Est-ce une raison suffisante pour renoncer à toute évaluation? Gage (1973, p. 342), qui, dans la recherche sur les comportements des maîtres, a souvent insisté sur la nécessité de n'apprécier l'enseignement qu'à travers ses fruits, et n'est donc pas

suspect de subjectivisme  
« Nous ne devrions pas nous inquiéter des enseignants qui reflètent nos préjugés. L'absence de réussite des élèves soit due à la compétence des enseignants soit à d'autres facteurs. »

Assurément, les valeurs prises en compte au moment de l'évaluation poursuivent (les deux étant ici être laissé seul juge : explicitement ou implicitement en poste. Exactement ce programme fondamental de devoir de l'éclairer et de

La difficulté de la méthode et de développement de la méthode présentée appellent aussi

La psychologie piagétienne d'utiles points de repère pour une analyse des réquisits de tout apprentissage imposé

Mialaret (1974) s'est intéressé à la difficulté formelle des contenus du volume du présent *Traité* publié une étude sur le test de l'instrument le plus facile pour la notation des notes de cours, des tests de Henry (1975) a proposé la méthode spécialement conçues pour

L'analyse lexicale paraît être intéressante. Il semble que dans le discours pédagogique, pour les aspects du rendement, pour un test.

La méthode d'analyse de la leçon et certains dictionnaires des mots fréquents (rendement, beaucoup...), des chaînes de propos...), des chaînons de problèmes (com

Outre les travaux de R. Hillier, ceux de J. Hillier, G. Fisher et d'autres des recherches, on ne peut

personnalité, de l'appren-  
méthodologie de l'ensei-  
ner. Certaines exigences  
alité, tenue régulière des  
catégorie des conditions.

nduire un train fou, ou  
entissages, un enseignant  
e son action, qu'il pour-  
ctifs très spécifiques. La  
en éducation constitue  
e plus progressé au cours  
lsheere, 1975).

édagogique, les objectifs  
d'une idéologie unique,  
les enseignants adhèrent  
ocèdent pas moins à une  
consciente.

un enseignant privilégie  
luer un enseignant dans  
, ne revêt donc guère de  
s'intéresse qu'à certains  
ns ce cadre, ou bien il  
nant auquel il a affaire  
fs précis que le professeur  
ne pas être d'accord sur  
plus d'évaluation d'effica-  
quation entre les fins, les

— On peut enseigner  
humainement mutilantes  
nent pour évident que la  
e n'est pas sûr. En outre,  
ances ou de complexité  
eurs devient de plus en  
notamment, seule une

eurs et des attitudes, les  
la validité devient indé-  
suffisante pour renoncer  
ui, dans la recherche sur  
nsisté sur la nécessité de  
fruits, et n'est donc pas

suspect de subjectivisme débridé, prend une position sans ambiguïté :

« Nous ne devrions pas hésiter à évaluer les comportements d'ensei-  
gnants qui reflètent nos valeurs, que leur corrélation avec la perfor-  
mance des élèves soit démontrée ou non. Si nous désirons que nos  
enseignants soient décents et polis, nous devons les juger à ces  
égards. »

Assurément, les valeurs épousées par les enseignants doivent être  
prises en compte au même titre que les objectifs cognitifs qu'ils  
poursuivent (les deux étant d'ailleurs liés). Mais l'éducateur ne peut  
ici être laissé seul juge : la société, la famille, l'étudiant lui confient  
explicitement ou implicitement une mission qu'il accepte en restant  
en poste. Exactement comme un architecte doit respecter le pro-  
gramme fondamental de son client, même s'il a aussi le droit et le  
devoir de l'éclairer et de le guider.

La difficulté de la matière par rapport au niveau de connaissances  
et de développement des élèves, ou le langage dans lequel elle est  
présentée appellent aussi une évaluation.

La psychologie piagétienne, en particulier, apporte à cet égard  
d'utiles points de repère développementaux. Plus particulièrement,  
une analyse des réquisits et un prétest devraient précéder l'abord de  
tout apprentissage important.

Mialaret (1974) s'est penché sur le problème de la densité et de  
la difficulté formelle des messages pédagogiques, dans le quatrième  
volume du présent *Traité*. Simultanément, De Landsheere (1973) a  
publié une étude sur le test de closure de W. Taylor, probablement  
l'instrument le plus facile pour évaluer la lisibilité ou l'intelligibilité  
des notes de cours, des textes employés et des manuels. Par la suite,  
Henry (1975) a proposé la première batterie de formules de lisibilité  
spécialement conçues pour la langue française.

L'analyse lexicale par ordinateur ouvre aussi des perspectives  
intéressantes. Il semble que la fréquence de certains mots, dans le  
discours pédagogique, peut être mise en relation avec certains  
aspects du rendement, par exemple la compréhension mesurée par  
un test.

La méthode d'analyse consiste en une comparaison automatique  
entre la leçon et certains dictionnaires (mis en mémoire d'ordinateur) :  
dictionnaire des mots fréquents, des mots vagues (presque, générale-  
ment, beaucoup...), des indices d'insistance sur les détails (noms  
propres...), des chaînons explicatifs (parce que, afin que...); des  
indices de problèmes (conflits, divergences...).

Outre les travaux de Rosenshine (1968), on retiendra, par exemple,  
ceux de J. Hiller, G. Fisher et W. Kaess (1969). Dans l'état actuel  
des recherches, on ne peut toutefois que constater l'apparition de

corrélations significatives et orientées dans le même sens, dans des situations comparables.

4° *Les apprentissages réalisés par les élèves.* — Ce que les élèves apprennent constitue la raison d'être de l'enseignement et donc le principal critère d'appréciation. Le nœud de la validité se trouve ici. On sait combien l'évaluation des produits de l'enseignement est difficile — nous allons y revenir; on ne peut cependant la négliger entièrement.

#### *Evaluation normative*

1° *Comparer à conditions égales.* — Si, au lieu d'évaluer l'enseignant en lui-même, on ambitionne de reconnaître les mérites relatifs de chacun et d'établir un classement de compétence professionnelle, le problème prend des dimensions supplémentaires.

Ou bien on continue à tenir compte du type d'enseignant et des objectifs spécifiques de chacun (mais comment trouver un dénominateur commun ?), ou bien on place tous les intéressés dans les mêmes conditions. Cette standardisation n'est pas nécessairement mutilante, car on peut varier les conditions (nombre d'élèves, âge, objectifs...), de façon à offrir à chacun la possibilité de manifester ses capacités et ses habiletés particulières. Des plans expérimentaux, aujourd'hui familiers à tous les chercheurs, permettent des rotations contrôlées apportant un maximum de garanties d'égalisation de chances et de traitements à tous les intéressés.

2° *Pouvoir discriminatif.* — L'exigence serait trop évidente pour qu'on la rappelât s'il ne fallait insister sur l'intérêt d'une discrimination multidimensionnelle. Sauf dans quelques cas exceptionnels, on ne peut se satisfaire de dichotomiser (efficaces-inefficaces) ou de trichotomiser (bons-moyens-mauvais) l'évaluation. Le processus d'enseignement est trop complexe pour autoriser des appréciations simplistes, manichéennes. Des profils reflètent mieux, en général, la réalité observée.

#### 2 | LA FIDÉLITÉ

On le sait, la validité et la fidélité ont partie liée : si, placé plusieurs fois consécutives devant une même réalité et s'intéressant aux mêmes aspects, l'évaluateur arrive à des conclusions fondamentalement différentes, rien ne permet plus de dire laquelle des évaluations est

valide. Ce problème est le même qui existe entre le tenant compte et la fidélité.

La méthode des incidences des enseignants montre que, l'observateur informé par des méthodes pédagogiques que de ce

Sans nier pour autant l'importance des éducateurs de grande expérience et à des techniques d'appréciation portées.

On distingue les méthodes qu'elles appellent. Elle est souvent automatisée (exemple : pose une question d'appréciation réduites à un

Captive l'attention

Les faiblesses des échelles sont bien connues et reconnues. Il existe aussi des comparaisons de fois où le professeur n'est pas évalué et des évaluations à inférences de façon opérationnelle et sur l'échelle, l'inférence peut n'en est pas entièrement précise. Les consignes nécessaires à la compréhension de De Bal, De Landsherre et De Landsherre ont été un problème en programmant l'évaluation.

Deuxième remarque : les précieuses informations que nous donnent comment les étudiants, les professeurs méritent certaines

L'important, en fin de compte

(1) MIALARET (1971) a soulevé un problème de validité dans une situation scolaire. Une évaluation a été faite sur un échantillon de 48 élèves et un maximum de 48 %.

DE LANDSHEERE et BAYER (1971) ont étudié la fidélité de quatre séries de leçons faites en première année de collège. La moyenne, toutes les quatre séries

même sens, dans des

— Ce que les élèves enseignent et donc le validité se trouve ici. de l'enseignement est cependant la négliger

d'évaluer l'enseignant les mérites relatifs de l'enseignement professionnelle, les résultats.

type d'enseignant et des méthodes pour trouver un dénominateur commun intéressés dans les mêmes critères. L'évaluation est nécessairement mutilante, elle ne mesure pas (élèves, âge, objectifs...), elle ne permet pas de manifester ses capacités intellectuelles, aujourd'hui elle mesure les rotations contrôlées, la probabilité de chances et de

est trop évidente pour faire l'objet d'un intérêt d'une discrimination dans les cas exceptionnels, on ne mesure pas les succès-inefficaces) ou de la validité. Le processus d'enseignement est donc des appréciations simplifiées, en général, la

est liée : si, placé plusieurs fois dans des situations intéressantes aux mêmes conditions fondamentales, les résultats des évaluations est

valide. Ce problème est bien connu, de même que l'espèce d'antagonisme qui existe entre le nombre de nuances dont une appréciation tient compte et la fidélité <sup>(1)</sup>.

La méthode des incidents critiques appliquée à l'évaluation des enseignants montre que, dans bien des cas, les remarques faites par l'observateur informent plus de sa personnalité et de ses conceptions pédagogiques que de ce qui se passe réellement en classe.

Sans nier pour autant toute validité aux jugements portés par des éducateurs de grande expérience, il importe donc de recourir à des méthodes et à des techniques permettant d'objectiver, en partie, les appréciations portées.

On distingue les méthodes d'évaluation selon le degré d'inférence qu'elles appellent. Elle est nulle dans le cas de comptages (éventuellement automatisés) de comportements strictement définis d'avance (exemple : pose une question); elle est maximale dans des échelles d'appréciation réduites à un continuum reliant deux pôles abstraits :

Captive l'attention

Ne captive pas l'attention

Les faiblesses des échelles d'évaluation sont aujourd'hui généralement connues et reconnues. Toutefois, deux remarques s'imposent. Il existe aussi des comptages à haute inférence (relever le nombre de fois où le professeur n'adapte pas son langage au niveau des élèves) et des évaluations à inférence très basse. Dans les échelles décrivant de façon opérationnelle et le trait à observer et chacun des degrés de l'échelle, l'inférence peut devenir pratiquement nulle. Le problème n'en est pas entièrement résolu pour autant. Car l'accumulation des précisions nécessaires à l'élimination des ambiguïtés conduit à des consignes dont la complexité devient à son tour une source d'erreurs. De Bal, De Landsherre et Paquay-Beckers (1977) ont résolu ce problème en programmant l'évaluation.

Deuxième remarque : la réaction *subjective* peut aussi apporter de précieuses informations et donc être recherchée comme telle. Savoir comment les étudiants, les inspecteurs, les directeurs perçoivent un professeur mérite certainement l'attention.

L'important, en fin de compte, est de savoir ce qu'on fait et de

<sup>(1)</sup> MIALARET (1971) a soumis à douze professeurs un film de quelques minutes présentant une situation scolaire. Une analyse du contenu avait révélé 52 éléments différents. Le pourcentage moyen d'éléments retenus est de 32,5 %, avec un minimum de 13,4 % et un maximum de 48 %. Pour un second film, la moyenne est de 27,2 %.

DE LANDSHEERE et BAYER (1969) constatent qu'une analyse encore relativement grossière de leçons faites en première année élémentaire appelle une nouvelle observation, en moyenne, toutes les quatre secondes.

ne pas accorder valeur de mesure rigoureuse à des réactions affectives vagues. Réagissant contre le néo-romantisme qui a déferlé ces derniers temps sur la pédagogie, Rosenshine (1971, p. 79) écrit :

« De même que certains prétendent reconnaître le grand art quand ils le voient, de même un nombre croissant d'éducateurs prétendent connaître la bonne éducation quand ils voient des événements correspondant à leurs intuitions poétiques. Les intuitions poétiques, jadis considérées comme points de départ de recherches, sont peu à peu devenues des conclusions suffisantes. »

### III

#### Les méthodes et les techniques

Dans une classification proposée en 1960 et toujours aussi éclairante, Mitzel distingue :

- l'étude des *caractéristiques* de l'enseignant permettant de *présager* qu'il s'acquittera efficacement de ses fonctions;
- l'étude descriptive du *processus* d'enseignement en vue du repérage de certains comportements ou de certains types d'interactions maître-élève(s) qui seraient en relation avec les apprentissages;
- l'étude des *produits* de l'enseignement, c'est-à-dire l'évaluation des apprentissages réalisés par les élèves.

#### I | L'ÉTUDE DES CARACTÉRISTIQUES

Les pronostics fondés sur des tests d'intelligence ou de personnalité continuent à s'avérer décevants. Qu'il existe une relation entre les aptitudes et la personnalité d'un enseignant, et les résultats qu'il obtient paraît indubitable, mais, la typologie adoptée d'emblée le rappelle, tel éducateur joue, avec plus ou moins de succès, sur la corde cognitive tandis que tel autre travaille dans un registre plus affectif.

Le modèle de Guilford permet d'expliquer aisément que, selon leurs constellations d'aptitudes, les éducateurs trouvent leur point fort dans le domaine pratique, symbolique, verbal ou social, et « opèrent » à des niveaux cognitifs différents. Il n'est pas rare que l'on

reproche aux enseignants de ne pas reconnaître toutes les aptitudes possibles de leurs élèves.

Fattu, cité par Flancon (1973), a étudié l'objectivité de relations entre l'enseignant et un enseignement efficace. Le nombre d'années d'expérience de l'enseignant, son sexe, son état civil, ses connaissances en sciences, son degré de confiance en soi, son diction un peu plus sûre, son attitude en relation avec l'observation de la classe, etc. E. Leclercq-Boxus (1973) a étudié la relation entre l'intensité de la guidance et le succès. A la liste de Fattu, on ajoute l'âge, l'état physique, de nombreux autres facteurs.

Evans (1959), dont les travaux ont été repris par d'autres auteurs, ne trouve pas de relation entre le degré de réussite dans l'enseignement et la réussite professionnelle.

En prenant quelque vingt caractéristiques suivies par les enseignants, comme de bon augure

- capacité à résoudre des situations d'enseignement;
- attitude démocratique;
- ordre et bonne planification;
- chaleur humaine dans la classe.

*La sélection avant l'entrée dans la profession.* Certaines caractéristiques de l'enseignant sont plus liées à la réussite que d'autres. Nous venons de voir que la maîtrise de la matière est une caractéristique importante.

La riche expérience de la recherche en pédagogie a permis de découvrir certaines recherches qui l'ont accompagnée. Une idée assez juste du rôle de la sélection à l'entrée dans la profession a été avancée. Trois exemples permettent de voir comment les connaissances sont liées à la réussite.

Pendant vingt ans (1950-1970), on a étudié la relation entre la réussite et la rédaction (tests narratif et descriptif) et la réussite à des tests collectifs portant sur des connaissances générales. Le calcul des corrélations a montré que de ces épreuves et la note

à des réactions affectives  
qui a déferlé ces derniers  
(p. 79) écrit :

maître le grand art quand  
d'éducateurs prétendent  
t des événements corres-  
tutions poétiques, jadis  
nerches, sont peu à peu

niques

o et toujours aussi éclai-

et permettant de *présager*  
tions;

ment en vue du repérage  
ins types d'interactions  
vec les apprentissages;  
c'est-à-dire l'évaluation

gence ou de personnalité  
e une relation entre les  
t, et les résultats qu'il  
rie adoptée d'emblée le  
moins de succès, sur la  
e dans un registre plus

uer aisément que, selon  
urs trouvent leur point  
e, verbal ou social, et  
Il n'est pas rare que l'on

reproche aux enseignants, implicitement au moins, de ne pas posséder toutes les aptitudes possibles au degré le plus élevé...

Fattu, cité par Flanders (1969), estime qu'on n'a pas pu établir objectivement de relation générale entre les caractéristiques suivantes et un enseignement efficace : intelligence mesurée par les tests, âge, nombre d'années d'expérience professionnelle, origine sociale de l'enseignant, sexe, état civil, intérêts professionnels, aptitudes spéciales. Les connaissances en sciences de l'éducation permettraient une prédiction un peu plus sûre. Cette dernière remarque peut être mise en relation avec l'observation de Turner (1965) et de M. Detheux et E. Leclercq-Boxus (1973) qui ont constaté une relation significative entre l'intensité de la guidance et le rendement des élèves en lecture. A la liste de Fattu, on peut certainement ajouter aussi la beauté physique, de nombreux traits de personnalité et les connaissances.

Evans (1959), dont les résultats sont confirmés par de nombreux auteurs, ne trouve pas de relation significative générale entre le degré de réussite dans les études préparant à l'enseignement et la réussite professionnelle.

En prenant quelque recul vis-à-vis de la littérature, il semble que les caractéristiques suivantes sont le plus généralement reconnues comme de bon augure :

- capacité à résoudre des problèmes, à s'adapter à différentes situations d'enseignement;
- attitude démocratique en éducation;
- ordre et bonne planification de l'enseignement;
- chaleur humaine dans la relation avec les élèves, sens de l'humour.

*La sélection avant l'entrée en formation.* — Elle n'est possible que si certaines caractéristiques permettent de prédire l'habileté pédagogique. Nous venons de voir combien notre savoir est peu sûr en la matière.

La riche expérience pratique scandinave et les centaines de recherches qui l'ont accompagnée (Marklund et Gran, 1974) donnent une idée assez juste du faible degré d'efficacité des procédures de sélection à l'entrée dans les institutions de formation des enseignants. Trois exemples permettront de schématiser l'état actuel de nos connaissances.

Pendant vingt ans (1948-1968), les épreuves utilisées en Suède ont compris la rédaction d'une dissertation, une épreuve d'éloquence (tests narratif et descriptif), des activités libres avec des enfants, et des tests collectifs portant sur les attitudes envers la profession enseignante. Le calcul des corrélations (E. Malmquist, 1956) entre chacune de ces épreuves et la note de compétence professionnelle à la fin des

études pédagogiques révèle une très faible validité prédictive.

En Finlande, M. Koskenniemi (1965) a suivi 72 enseignants depuis leur sélection en vue de leur formation jusqu'à dix ans après le début de leur activité professionnelle. De nouveau, il ne trouve guère de corrélation entre les épreuves de sélection et la réussite professionnelle. « Un trait commun aux enseignants qui réussissaient le moins bien était leur manque de compréhension des enfants et leur manque d'aptitude à structurer une situation d'enseignement. »

P. Sundgren (1970) a cependant réussi à établir une relation entre les attitudes des élèves-maîtres pendant leur formation et l'évolution de leurs élèves, deux ans après l'obtention du diplôme d'enseignement.

Caractère de l'enseignant	Changement observé chez l'élève
1a. Attitude autoritaire	Grand progrès des connaissances
1b. Social, humain	Léger progrès des connaissances
2a. Social, humain, capable de s'effacer	Evolution sociale positive
2b. Tourné vers les résultats, manquant d'assurance sur le plan social	Evolution sociale négative
3a. Indépendant, tourné vers l'élève	Evolution positive vers une stabilité émotive
3b. Tourné vers les résultats, manquant d'assurance, grand besoin d'organisation	Evolution négative vers une instabilité émotive

Au terme de leur analyse des résultats scandinaves, Marklund et Gran concluent :

« Il n'existe pas d'« aptitude enseignante » simple et non équivoque, indépendante de la situation d'enseignement. L'« inaptitude » à l'enseignement semble plus facile à définir. La sélection pour les études conduisant à l'enseignement doit, par conséquent, avoir pour objectif principal d'éviter les échecs présumés. »

Nous partageons ce point de vue. Plus la recherche avance, et mieux on reconnaît le rôle capital que jouent les attitudes des enseignants dans le processus d'éducation. Comme ces attitudes ne peuvent s'observer que pendant des activités éducatives réelles, la sélection durable, en cours de formation, semble s'imposer. En particulier, des objectifs minimum, en matière d'attitudes, doivent être assignés aux élèves-maîtres. On refusera les étudiants qui n'atteignent pas ces objectifs avant la fin de leurs études, au cours desquelles des actions systématiques d'implantation ou de modification d'attitudes auraient été déployées.

## 2 | L'ÉTUDE DESCRIPTIVE DES PROCESSUS D'ENSEIGNEMENT

Rosenshine et Furst (1973) distinguent deux grandes méthodes d'enregistrement des observations :

## L'ÉVALUATION DES ENSEIGNANTS

- les systèmes de signes où pendant la période d'observation (mesure ordinaire);
- les systèmes de catégories enregistrées et où l'on recense (mesure nominale).

### Les systèmes de signes

Vraisemblablement, tous les systèmes d'enseignement se sont constitués à partir de certaines de ces méthodes. Les aspects que les chercheurs ont jugés importants; rarement, ils ont été structurés. En outre, plusieurs sont à haute inférence (l'inférence n'étant pas définie opérationnellement) de validité et de fidélité.

On recourt fréquemment à ces instruments. La plus grande difficulté n'est pas, à elle seule, une simple expression l'enseignant a décidé d'observer.

On sait qu'au terme de sa recherche Ryans (1960) conclut que, parmi les 6 000 enseignants, observés, on a ramené à trois dichotomies :

- aimable - distant, respectueux
- systématique - inorganisé
- imaginaire - sans imagination

Il est frappant de constater que ces données recueillies à l'aide de méthodes différentes sont si fortement corrélées. On a retravaillé par Mialaret (1973) et on trouverait pas le facteur imaginatif.

(1) En codage multiple, les méthodes sont indépendantes (ex. : qui partent de catégories et sont indépendantes).

(2) Nous ne revenons pas sur les effets de stabilité centrale, effets de stabilité.

GUILFORD (1954), REMMERS (1960).

le validité prédictive. a suivi 72 enseignants n jusqu'à dix ans après nouveau, il ne trouve sélection et la réussite gnants qui réussissaient hension des enfants et ation d'enseignement. » tablir une relation entre orformation et l'évolution iplôme d'enseignement.

observé chez l'élève

rés des connaissances

és des connaissances

ociale positive

ociale négative

ositive vers une stabilité émotive

égative vers une instabilité

andinaves, Marklund et

imple et non équivoque,

t. L' « inaptitude » à

La sélection pour les

conséquent, avoir pour

s. »

a recherche avance, et

les attitudes des ensei-

ces attitudes ne peuvent

ves réelles, la sélection

poser. En particulier,

s, doivent être assignés

qui n'atteignent pas ces

s desquelles des actions

on d'attitudes auraient

ENSEIGNEMENT

deux grandes méthodes

- les *systèmes de signes* où un événement n'est compté qu'une fois pendant la période d'observation et fait l'objet d'une *évaluation* (mesure ordinale);
- les *systèmes de catégories* où l'occurrence de chaque événement est enregistrée et où l'on recourt à un codage unique ou multiple <sup>(1)</sup> (mesure nominale).

### *Les systèmes de signes*

Vraisemblablement, tous les inspecteurs, tous les maîtres de stages d'enseignement se sont construit leur propre système de signes. Des centaines de ces systèmes ont été publiés. Le plus souvent, ils reflètent les aspects que les constructeurs tiennent personnellement pour importants; rarement, ils traduisent une théorie pédagogique très structurée. En outre, presque toutes les échelles d'évaluation utilisées sont à haute inférence (le trait à évaluer et les degrés de l'échelle n'étant pas définis opérationnellement). Ainsi s'explique le manque de validité et de fidélité si souvent dénoncé <sup>(2)</sup>.

On recourt fréquemment à l'analyse factorielle pour épurer les instruments. La plus grande rigueur ainsi obtenue ne constitue toutefois pas, à elle seule, une garantie de validité : on réduit à une plus simple expression l'ensemble des événements que l'on avait initialement décidé d'observer.

On sait qu'au terme d'une recherche souvent citée en exemple, Ryans (1960) conclut que le comportement fondamental de plus de 6 000 enseignants, observés dans 1 700 écoles différentes, peut être ramené à trois dichotomies :

- aimable - distant, réservé;
- systématique - inorganisé;
- imaginatif - sans inspiration.

Il est frappant de constater qu'une analyse factorielle des données recueillies à l'aide de l'échelle Stanford-ERENS (voir ci-après), instrument soigneusement mis au point à l'Université de Stanford, puis retravaillé par Mialaret (1971) et ses étudiants de Montréal, ne révélerait pas le facteur *imaginatif - sans inspiration* isolé par Ryans.

<sup>(1)</sup> En codage multiple, un même événement est codé selon plusieurs modalités indépendantes (ex. : qui parle; quel objectif est poursuivi; quelle matière est traitée) ou dépendantes : catégorie et sous-catégorie (ex. : organise; pose une question fermée).

<sup>(2)</sup> Nous ne revenons pas à la critique générale des échelles d'évaluation (tendance centrale, effets de stéréotype, effets de halo, etc.). Voir notamment à ce sujet GUILFORD (1954), REMMERS (1963), DE LANDSHEERE (1974).



		Impossible de noter	Nul	Très inférieur	Inférieur	Faible	Moyen	Fort	Supérieur	Très remarquable	Exceptionnel
<b>Objectifs</b>											
1. Exposé des objectifs	Le professeur annonce clairement ce qu'il va faire (il présente les buts ou les objectifs du cours).	0	1	2	3	4	5	6	7	8	9
2. Justesse des objectifs	Les buts que le professeur présente ne sont ni trop faciles, ni trop difficiles pour les étudiants. Ces buts conviennent à leur niveau et à leurs capacités.	0	1	2	3	4	5	6	7	8	9
<b>Préparation</b>											
3. Enchaînement du cours	Les liens entre les différentes parties du cours sont clairs et logiques. Le plan du cours permet aux étudiants de bien apprendre.	0	1	2	3	4	5	6	7	8	9
4. Choix du contenu	Le contenu du cours est bien adapté aux buts annoncés par le professeur et au niveau des étudiants.	0	1	2	3	4	5	6	7	8	9
5. Choix des moyens	Le professeur emploie une méthode, un matériel et des moyens d'enseignement bien adaptés au contenu de son cours.	0	1	2	3	4	5	6	7	8	9
<b>Exécution</b>											
6. Début du cours	Le professeur obtient rapidement l'attention des étudiants qui se mettent d'eux-mêmes au travail.	0	1	2	3	4	5	6	7	8	9
7. Clarté du cours	Le cours est clair et facile à comprendre. Le professeur présente plusieurs points de vue, si nécessaire. Il choisit bien ses exemples.	0	1	2	3	4	5	6	7	8	9
8. Rythme du cours	Avant de passer à une nouvelle partie, le professeur s'assure que tous ont bien compris. Il sait faire progresser son cours au rythme de sa classe. Il ne perd jamais de vue ses étudiants.	0	1	2	3	4	5	6	7	8	9
9. Participation et attention de l'étudiant	La classe est attentive. Les étudiants participent activement au cours.	0	1	2	3	4	5	6	7	8	9
10. Fin du cours	Le professeur met fin à son cours quand il constate que les étudiants ont atteint ce qu'il désirait. Il résume ce qui s'est dit et passé durant le cours.	0	1	2	3	4	5	6	7	8	9
11. Relations professeur-étudiants	Les relations entre le professeur et les étudiants sont harmonieuses.	0	1	2	3	4	5	6	7	8	9
<b>Contrôle</b>											
12. Variété des moyens de contrôle	Le professeur emploie plusieurs façons de contrôler ce que les étudiants ont appris. Dans cette évaluation, il tient compte de tous les buts qu'il s'était fixés.	0	1	2	3	4	5	6	7	8	9
13. Utilisation des résultats de contrôle	Afin d'améliorer leur rendement, le professeur et les étudiants analysent avec soin les résultats de cette évaluation.	0	1	2	3	4	5	6	7	8	9

(1) Traduction et adaptation du *Stanford Teacher Competence Appraisal Guide* par l'équipe de recherche de l'École normale supérieure (ERENS) de l'Université de Montréal.

Certains traits se prêtent à être les objets privilégiés de la presque totalité de l'occupation. La praticabilité de l'évaluation s'opère en particulier ceux exigent un enregistrement et une analyse des données.

Bref, les systèmes d'évaluation utilisés à l'avenir. Ils varient de l'évaluation globale, évaluations partielles. Il serait souhaitable d'effectuer — en particulier

- 1° d'utiliser les conclusions sans redondances et les
- 2° de dresser l'inventaire des traits d'économiser les moyens de construction.

Les traits suivants sont ceux des principaux auteurs de la liste. Ils visent à encourager la poursuite des objectifs, l'enthousiasme, imagination, bon sens, symptomatique qu'en analysant les résumés rédigés par des étudiants et Hildebrand et Wilsch. Ces traits : clair dans ses relations avec les étudiants, stimulante, enthousiaste.

Le système proposé est basé après une vingtaine d'années d'expérience complexe destiné au professeur.

*Echelles ancêtres par l'Université de Montréal* assez d'attention à la liste de Cantrill (1960) et présentée aux évaluateurs de l'évaluation idéal.

D'abord, l'évaluateur souhaite, soit en général, soit en particulière. Ensuite, on l'in-

	L'ÉVALUATION DES ENSEIGNANTS							
	Très inférieur	Inférieur	Faible	Moyen	Fort	Supérieur	Très remarquable	Exceptionnel
2	3	4	5	6	7	8	9	
2	3	4	5	6	7	8	9	
2	3	4	5	6	7	8	9	
2	3	4	5	6	7	8	9	
2	3	4	5	6	7	8	9	
2	3	4	5	6	7	8	9	
2	3	4	5	6	7	8	9	
2	3	4	5	6	7	8	9	
2	3	4	5	6	7	8	9	
2	3	4	5	6	7	8	9	
2	3	4	5	6	7	8	9	
2	3	4	5	6	7	8	9	
2	3	4	5	6	7	8	9	
2	3	4	5	6	7	8	9	

Certains traits se prêtent difficilement aux comptages et devraient être les objets privilégiés de systèmes de signes. On remarquera que la presque totalité de l'échelle Stanford-ERENS répond à cette préoccupation. La praticabilité constitue aussi un facteur décisif : une évaluation s'opère en peu de temps, tandis que des comptages rigoureux exigent un enregistrement, un codage, puis un traitement et une analyse des données.

Bref, les systèmes de signes continueront certainement à être utilisés à l'avenir. Ils varieront en fonction des besoins des utilisateurs : évaluation globale, évaluation d'une catégorie particulière de comportements. Il serait souhaitable d'exploiter les centaines de travaux effectués — en particulier :

- 1° d'utiliser les conclusions des analyses factorielles pour réduire les redondances et les confusions;
- 2° de dresser l'inventaire des échelles *descriptives opérationnelles* afin d'économiser les mois de travail souvent nécessaires à leur construction.

Les traits suivants sont généralement tenus pour importants par les principaux auteurs : accueillir les idées des étudiants et s'efforcer de les valoriser, encourager, bien planifier et organiser l'enseignement, poursuivre des objectifs clairs, conscience professionnelle, enthousiasme, imagination, bonnes relations avec les étudiants. Il est d'ailleurs symptomatique qu'en analysant un grand nombre de questionnaires rédigés par des étudiants pour évaluer leurs professeurs Eble (1970) et Hildebrand et Wilson (1970) retrouvent pratiquement cinq de ces traits : clair dans ses exposés et ses explications, suscite la discussion avec les étudiants, stimule la réflexion, s'intéresse aux étudiants, enthousiaste.

Le système proposé par Verbist et Van Herreweghe (1974), après une vingtaine d'années d'essais, fournit un exemple d'ensemble complexe destiné au perfectionnement de l'enseignement.

*Echelles ancrées par l'évaluateur.* — On n'a sans doute pas accordé assez d'attention à la technique proposée par Kilpatrick et Cantrill (1960) et présentée par H. Remmers (1963); elle permet aux évaluateurs de situer un enseignant par rapport à leur idéal.

D'abord, l'évaluateur décrit librement l'enseignant tel qu'il le souhaite, soit en général, soit à propos d'une caractéristique particulière. Ensuite, on l'invite à décrire l'enseignant le moins bon qu'il

puisse imaginer. Ayant ainsi obtenu les deux pôles de l'échelle, on demande alors d'évaluer un enseignant particulier.

10
9
8
7
6
5
4
3
2
1
0

#### Les systèmes de catégories

Les systèmes de catégories destinés à l'analyse des interactions maître-élèves se sont développés très rapidement au cours des quinze dernières années. Rosenshine et Furst (1973) en dénombrent 120 pour les pays anglo-saxons.

Les chercheurs s'influencent mutuellement et la distinction entre instruments analytiques et instruments synthétiques que nous avons proposée en 1970 semble perdre de plus en plus sa pertinence.

Rosenshine et Furst classent, selon la source des variables, 73 systèmes se rapportant directement à la pratique scolaire, contenus dans le compendium de Simon et Boyer (1967-1970) :

- a) systèmes construits à partir d'une théorie élaborée ou d'une recherche expérimentale (25) :
- référence explicite; exemple : Taba (Piaget-Bruner);
  - référence implicite; exemple : Flanders (psychologie de la relation inférieur-supérieur);
- b) modification ou synthèse de systèmes existants :
- développement d'un système (12); exemple : Amidon et Hunter à partir de Flanders;

— synthèse de plusieurs systèmes à partir de Ashner et al.

- c) catégories inventées par l'enseignant; se fonde sur une longue expérience.

Pour les pays de langue anglaise, on compte des conditions opératoires affectives inspirées de la méthode de Landsheere et al. Le système de Hughes en Suisse (Postic (1971) utilise de nombreuses catégories. De Landsheere-Bayer, et pour les actes pédagogiques du professeur, compte des comportements de systèmes de De Landsheere et al. de Bloom. Dussault *et al.* ont développé le système de Joyce (1967).

En prenant du recul, on peut dire que le plus employé de tous les systèmes est astucieux et mal adapté aux sources. De Landsheere (1967) a certaines catégories et a été adopté par les anglo-saxons, les autres systèmes de M. Hughes (1959) et de Smith et Meux (1962) ont une fréquence des références élevées. On indique une large dissémination de la formation des maîtres.

On a reproché à ces systèmes, croyons-nous, à tort :

1. En raison même de la complexité des enseignants s'est trouvée réduite l'importance de reprendre fréquemment les mêmes catégories mentales sûres, permettant de contrôler les comportements conformes et aussi de faire ressortir les lacunes.

(<sup>1</sup>) Les principaux systèmes de catégories ont été étudiés par E. BAYER, *Comment les maîtres enseignent*, ministère de l'Éducation nationale, Paris, 1967, en langue française (sauf BAYER, et G. MIALARET, *Traité des sciences de l'éducation*, p. 166 sq. Voir aussi M. POSTIC

pôles de l'échelle, on  
lier.

— synthèse de plusieurs systèmes (12); exemple : Amidon, à partir de Ashner et Gallagher, Flanders, Hughes, Taba;

c) catégories inventées par les auteurs (24); exemple : Hughes, qui se fonde sur une longue pratique de l'inspection scolaire.

Pour les pays de langue française, Ferry *et al.* (1968) tiennent compte des conditions opératoires de la communication et de conditions affectives inspirées de la psychologie de la relation inférieur-supérieur. De Landsheere et Bayer (1969) restructurent et développent le système de Hughes en se référant à la psychologie néo-behavioriste. Postic (1971) utilise de nombreuses catégories des systèmes Hughes, De Landsheere-Bayer, et propose une grille à double entrée combinant les actes pédagogiques du professeur et leur intention dominante; il tient compte des comportements non verbaux (1). Bayer (1973) combine les systèmes de De Landsheere-Bayer, de Bellack et la taxonomie cognitive de Bloom. Dussault *et al.* (1973) ont réalisé une adaptation française du système de Joyce (1967), axé sur la flexibilité de l'enseignement.

En prenant du recul, on constate que quelques systèmes émergent. Le plus employé de tous est certainement celui de Flanders. Le système est astucieux et Mialaret (1974) en a rappelé les vastes ressources. De Landsheere (1973) a toutefois démontré l'ambiguïté de certaines catégories et a proposé une grille clarifiée. Dans les pays anglo-saxons, les autres systèmes qui exercent le plus d'influence sont ceux de M. Hughes (1959), de H. Taba (1965), de A. Bellack (1965), de Smith et Meux (1964). Dans les pays de langue française, la fréquence des références au système De Landsheere-Bayer semble indiquer une large dissémination dans les milieux de recherche et de formation des maîtres.

On a reproché à ces systèmes leur caractère descriptif. C'est, croyons-nous, à tort :

1. En raison même de l'impasse où la recherche sur l'évaluation des enseignants s'est trouvée après la seconde guerre mondiale, il importait de reprendre fondamentalement le problème.

2. Pareils systèmes, élaborés sur des bases théoriques et expérimentales sûres, permettent de reconnaître, en cours d'enseignement, les comportements conformes aux options méthodologiques prises et aussi de faire ressortir les lacunes ou les contradictions. Ainsi s'explique

(1) Les principaux systèmes américains sont présentés dans G. DE LANDSHEERE et E. BAYER, *Comment les maîtres enseignent. Analyse des interactions verbales en classe*, Bruxelles, ministère de l'Éducation nationale, Organisation des Études, 1975, 3<sup>e</sup> éd. Les systèmes en langue française (sauf BAYER, 1973) sont synthétisés par G. MIALARET, in M. DEBESSE et G. MIALARET, *Traité des sciences pédagogiques*, 4 : *Psychologie de l'éducation*, Paris, PUF, 1974, p. 166 sq. Voir aussi M. POSTIC (1977).

analyse des interactions  
nt au cours des quinze  
en dénombrent 120

et la distinction entre  
iques que nous avons  
us sa pertinence.

e des variables, 73 sys-  
que scolaire, contenus  
-1970) :

ie élaborée ou d'une

piaget-Bruner);

rs (psychologie de la

tants :

xemple : Amidon et

d'ailleurs que les systèmes de catégories sont de plus en plus utilisés en micro-enseignement et à des fins de diagnostic et de perfectionnement.

3. On s'aperçoit aujourd'hui que ces systèmes peuvent être employés pour repérer des maîtres pratiquant le même style d'enseignement ou des styles contrastés, et on obtient ainsi des variables indépendantes pour des recherches expérimentales. Nous allons voir, d'ailleurs, que le système de Flanders a notablement contribué à l'avancement de la recherche sur l'évaluation des produits.

La majorité des systèmes de catégories portent sur les comportements verbaux de l'enseignant. La nette prédominance quantitative de ces comportements a convaincu de nombreux chercheurs que cette forme de communication est représentative de l'ensemble des interactions maître-élèves.

Périodiquement, on objecte que l'implicite, le non-verbal exercent une influence insuffisamment étudiée (Léon, 1975). La recherche expérimentale est, en effet, pauvre en la matière. C'est d'autant plus regrettable que certains incidents critiques et aussi certains renforcements, probablement fort importants, se situent au niveau du geste et de la mimique.

Ainsi apparaît l'importance de la sémiologie du geste pédagogique proposée par M. Fauquet et S. Strasfogel (1972). Ils distinguent trois grandes catégories gestuelles : les gestes d'organisation (gestes d'action), les gestes de sanction, les gestes symboliques et les gestes liés au comportement verbal (gestes mimétiques). Selon les personnalités et le niveau d'habileté professionnelle, les rapports entre les différentes catégories de gestes varient. Chez les débutants, les gestes d'organisation sont les plus nombreux.

Une piste de recherche est également apportée par B. Rosenshine (1968). Etudiant les caractéristiques comportementales d'enseignants qui expliquent très clairement une même matière, et d'autres reconnus comme les plus confus, il retient trois prédicteurs d'efficacité, dont la quantité de gestes et de mouvements faits par les professeurs.

A. Delchambre (1975) puis De Landsheere et Delchambre ont entrepris une recherche approfondie sur l'incidence des comportements non verbaux. Un ensemble de leçons ont été enregistrées sur magnétoscope, puis codées selon le système d'analyse De Landsheere-Bayer. Les comportements non verbaux, interprétés par des observateurs bien entraînés (concordance supérieure à .85), ont été situés par rapport aux interactions verbales.

Les résultats de cette recherche, toujours en cours au Laboratoire de Liège, semblent confirmer l'hypothèse selon laquelle il existerait une forte association entre la signification du comportement verbal et du non-verbal.

Les points suivants re

- 1° Pris dans leur totalité sont beaucoup plus
  - 2° L'apport essentiel de situerait dans le dom
  - 3° L'analyse et le relevé et d'affectivité négati
  - 4° Les fonctions d'affect
  - 5° Les interventions sil
- 1 % de l'ensemble des

### 3 | L'ÉVALUATION DES E

*Bilan.* — L'évaluation de grandes difficultés.

Alors que les effets d' et à long terme, seuls le peuvent être sûrement a d'une leçon, et pour aut aient été mesurées, il est trop de risques d'erreur tard, la part qui revient à quement plus décelable a vraiment dans l'éducation Qui oserait prétendre qu de quelques mois d'ense critique, de la créativité, de la capacité d'analyse e électivement des effets heu l'enseignement de plusieurs dans l'habileté technique (essentiellement mémoire ment — ce qui explique p presque toujours à ce niv du bon enseignement.

Ainsi s'explique que, cinquante dernières années d'observations suivies de

de plus en plus utilisés en et de perfectionnement. Les systèmes peuvent être de même style d'enseignement ainsi des variables fondamentales. Nous allons voir, cela a contribué à la production des produits.

Il ressort sur les comportements la dominance quantitative des chercheurs que cette méthode de l'ensemble des inter-

actions, le non-verbal exercent (Landsheere, 1975). La recherche montre que c'est d'autant plus vrai que certains renforcent au niveau du geste

du geste pédagogique (Landsheere, 1972). Ils distinguent trois types de gestes (gestes d'action), les gestes liés au contenu. Selon les personnalités des enseignants, les différents types de gestes, les gestes d'organisa-

tion rapportée par B. Rosenfeld. Les comportements d'enseignement de même matière, et d'autres facteurs prédictifs d'efficacité, ont été étudiés par les professeurs. Les faits par les professeurs. Rosenfeld et Delchambre ont étudié l'incidence des comportements ont été enregistrées sur l'analyse De Landsheere (1975), ont été situés par

un cours au Laboratoire de psychologie sur laquelle il existerait un comportement verbal

Les points suivants ressortent de l'étude de Delchambre :

- 1° Pris dans leur totalité, les comportements non verbaux du maître sont beaucoup plus nombreux que ses comportements verbaux ;
- 2° L'apport essentiel de l'analyse des comportements non verbaux se situerait dans le domaine du *feedback* (évaluation positive, négative ; fonctions d'affectivité positive, négative) ;
- 3° L'analyse et le relevé des fonctions verbales d'affectivité positive et d'affectivité négative permettent de prédire les proportions respectives des comportements d'affectivité non verbaux ;
- 4° Les fonctions d'affectivité négative sont exprimées verbalement, tandis que les fonctions d'affectivité positive sont plus fréquemment émises non verbalement ;
- 5° Les interventions silencieuses d'affectivité représentent environ 11 % de l'ensemble des fonctions verbales relevées pendant la leçon.

### 3 | L'ÉVALUATION DES PRODUITS

*Bilan.* — L'évaluation des produits de l'enseignement se heurte à de grandes difficultés.

Alors que les effets d'un enseignement s'exercent à court, à moyen et à long terme, seuls les effets à court, voire à très court terme, peuvent être sûrement attribués à l'action des maîtres : à la fin d'une leçon, et pour autant que les connaissances existant au départ aient été mesurées, il est possible de mesurer un nouvel acquis, sans trop de risques d'erreur. Quelques mois ou quelques années plus tard, la part qui revient à l'action de tel ou tel professeur n'est pratiquement plus décelable avec certitude. Pourtant, tout ce qui compte vraiment dans l'éducation ne se construit, ne s'installe qu'avec lenteur. Qui oserait prétendre quantifier l'influence de quelques semaines ou de quelques mois d'enseignement sur le développement de l'esprit critique, de la créativité, de la sensibilité linguistique, du sens social, de la capacité d'analyse et de synthèse ? Et d'ailleurs, à qui attribuer électivement des effets heureux quand l'étudiant reçoit simultanément l'enseignement de plusieurs maîtres ? Le progrès dans le dressage, dans l'habileté technique, dans les activités cognitives inférieures (essentiellement mémoire et productions convergentes), se mesure aisément — ce qui explique pourquoi les examens traditionnels se situent presque toujours à ce niveau —, mais il n'est pas le critère supérieur du bon enseignement.

Ainsi s'explique que, parmi les milliers de recherches de ces cinquante dernières années, on ne trouve guère plus d'une soixantaine d'observations suivies de calcul de corrélations, ou d'expériences dans

lesquelles les apprentissages des élèves sont pris pour critères. Et presque toutes ces études datent de 1966 ou après (Rosenshine, 1971). Sur mille recherches recensées par Burhart (1969) en 1967, vingt seulement prennent le développement (*growth*) de l'élève comme critère.

Parmi ces recherches, il importerait de distinguer, d'une part, les expériences de laboratoire et les expériences sur le terrain et, d'autre part, les expériences où l'enseignement même est peu contrôlé, et les expériences où les variables d'enseignement sont systématiquement manipulées.

Il est clair qu'en passant du laboratoire sur le terrain, on rencontre une telle quantité de variables situationnelles nouvelles que les résultats initialement acquis peuvent être presque tous remis en question (Stake, 1974). A notre connaissance, il n'existe aucune recherche approfondie sur la différenciation qui s'opère à mesure qu'un ensemble méthodologique mis au point en laboratoire se dissémine dans la pratique.

Toutefois, les observations suivantes donnent une première idée de ce que l'on rencontrerait vraisemblablement. Dressant le bilan de 27 recherches sur l'enseignement de la lecture en première année primaire (187 classes), Bond et Dykstra (1967) notent que, pour une même méthode générale, les résultats diffèrent beaucoup et concluent que ce sont les maîtres, les activités réelles et les situations qu'il importe d'étudier. Dans une situation beaucoup mieux contrôlée où tous les professeurs enseignent la même unité de biologie (mêmes objectifs, même matériel d'enseignement et de testing, méthodologie bien définie, groupes d'élèves équivalents), Gallagher (1966, 1970) observe des différences significatives chez six enseignants observés. L'un parle pendant 66 % et un autre pendant 95 % du temps; ils travaillent à des niveaux cognitifs très différents.

Beaucoup de recherches de ces dix dernières années utilisent la matrice de Flanders pour analyser la façon d'enseigner et étudier les corrélations avec certains comportements observés chez les élèves. Flanders (1969) a dressé lui-même le bilan de ces recherches. Pour la première fois peut-être, une concordance dans les résultats expérimentaux apparaît. On peut maintenant affirmer, estime Flanders (p. 1426), « que le pourcentage d'interventions verbales du professeur qui utilisent des idées ou des opinions antérieurement exprimées par les élèves est en relation directe avec les attitudes vis-à-vis de l'enseignement et des études, ainsi qu'avec les performances des élèves, calculées en tenant compte de leurs capacités initiales ».

Flanders était cette affirmation par des résultats obtenus soit en laissant les maîtres pratiquer leur style d'enseignement habituel

(Flanders; Morrison; L... spécialement à pratiquer... et Flanders; Schantz; Fil...

Ces résultats rejoignent... shine (1971) qui, à la dem... la littérature de recherche... apprentissages des élèves...

Les quelques résultats... la recherche n'en restent... analyses sont, en général... d'attitudes ne le sont gu...

Les conditions expéri... blement raffinées pour de... de la spécificité d'efficaci...

Comme le souligne Po... latives restent « ambiguës... lées (...). Est-ce qu'un co... l'élève à de meilleures p... deux variables sont-elles... facteurs » ?

Nous avons remarqué... être justement comparés... enseignement, que si les c... semblables. McNeil et Po... rale suivante :

1. On recherche des ob... et pour les élèves.
2. On informe les maî... pour établir si les objectif...
3. On constitue des gr... dans une même populatio... on ne retient que les élèv...
4. Il arrive que l'on... compris des manuels.
5. Différents maîtres e... de leur choix, pour atteind... possible, on pratique une... à contrôler le maximum d...
6. A la fin de chaque a... ou psychomoteurs sont c... inconnus des maîtres. Pa... rétention après une périod...

pris pour critères. Et  
près (Rosenshine, 1971).  
(1969) en 1967, vingt  
*eth*) de l'élève comme

distinguer, d'une part,  
ences sur le terrain et,  
même est peu contrôlé,  
ment sont systématique-

le terrain, on rencontre  
elles nouvelles que les  
presque tous remis en  
nce, il n'existe aucune  
qui s'opère à mesure  
oint en laboratoire se

ment une première idée  
ment. Dressant le bilan  
cture en première année  
) notent que, pour une  
t beaucoup et concluent  
s et les situations qu'il  
ucoup mieux contrôlée  
mité de biologie (mêmes  
le testing, méthodologie  
Gallagher (1966, 1970)  
x enseignants observés.  
ant 95 % du temps; ils  
ents.

nières années utilisent la  
'enseigner et étudier les  
observés chez les élèves.  
ces recherches. Pour la  
ans les résultats expéri-  
firmer, estime Flanders  
s verbales du professeur  
eurement exprimées par  
udes vis-à-vis de l'ensei-  
performances des élèves,  
és initiales ».

résultats obtenus soit en  
'enseignement habituel

(Flanders; Morrison; Lashier; Pankratz), soit en les entraînant spécialement à pratiquer le style direct ou le style indirect (Amidon et Flanders; Schantz; Filson; Miller; Reed).

Ces résultats rejoignent les indications fournies par Rosenshine (1971) qui, à la demande de l'IEA, a systématiquement dépouillé la littérature de recherche pour en extraire les travaux prenant les apprentissages des élèves comme variable dépendante.

Les quelques résultats encourageants récoltés jusqu'à présent par la recherche n'en restent pas moins doublement insatisfaisants : les analyses sont, en général, grossières et les mesures de performance ou d'attitudes ne le sont guère moins.

Les conditions expérimentales et les mesures doivent être inlassablement raffinées pour découvrir les facteurs explicatifs du degré et de la spécificité d'efficacité d'un enseignant.

Comme le souligne Postlethwaite (1974, p. 20), les études corrélatives restent « ambiguës quant à l'aspect causal des relations révélées (...). Est-ce qu'un comportement approuvateur du maître incite l'élève à de meilleures performances ou en résulte ? Ou bien ces deux variables sont-elles déterminées par une troisième série de facteurs » ?

Nous avons remarqué déjà que différents professeurs ne peuvent être justement comparés entre eux, en fonction du produit de leur enseignement, que si les conditions dans lesquelles ils travaillent sont semblables. McNeil et Popham (1973) proposent la démarche générale suivante :

1. On recherche des objectifs d'enseignement neufs pour les maîtres et pour les élèves.
2. On informe les maîtres de certaines mesures qui seront opérées pour établir si les objectifs sont atteints par les élèves.
3. On constitue des groupes équivalents d'élèves par tirage au sort dans une même population et, si certains réquisits sont nécessaires, on ne retient que les élèves qui les possèdent.
4. Il arrive que l'on fournisse aussi un matériel didactique, y compris des manuels.
5. Différents maîtres entreprennent d'enseigner, selon la méthode de leur choix, pour atteindre les mêmes objectifs. Dans la mesure du possible, on pratique une rotation dans les groupes d'élèves, de façon à contrôler le maximum de variables.
6. A la fin de chaque activité, les apprentissages cognitifs, affectifs ou psychomoteurs sont contrôlés à l'aide d'instruments en partie inconnus des maîtres. Parfois, on procède aussi à un contrôle de rétention après une période donnée.



7. On constate généralement, au terme des expériences, que certains enseignants obtiennent presque toujours de meilleurs résultats que d'autres. Ce sont les comportements de ces individus qu'il importe d'étudier par opposition aux autres.

Réussira-t-on ainsi à évaluer un jour l'influence d'un enseignant particulier sur des apprentissages à long terme ? D'une façon certaine, il ne semble pas. Pas plus qu'on ne pourrait démontrer dans quelle mesure la bonne santé d'un homme, en général, et sa longévité sont attribuables au médecin de famille. A partir de cette constatation, Gage (1972, p. 177) prend une position marquée au coin du bon sens. Les médecins tirent argument des disciplines de base et de leur savoir empirique en faveur de telle ou telle pratique qui, d'ailleurs, entraîne souvent une amélioration immédiate. Pourquoi ne procéderait-on pas de même dans l'enseignement à partir de disciplines comme la logique, les théories de l'apprentissage, les règles d'hygiène mentale ? De plus, ne vaut-il pas mieux mesurer correctement les effets à court terme que de ne rien mesurer du tout ?

*Le système contractuel.* — Issu du vaste mouvement de recherche contemporain sur la définition opérationnelle des objectifs de l'éducation, le système contractuel ouvre des perspectives d'un grand intérêt.

Les caractéristiques du système (dont il existe déjà de nombreuses variantes) sont clairement cernées par McNeil et Popham (1973, p. 234) dont nous nous inspirons directement :

1° L'enseignant et l'inspecteur se mettent d'abord d'accord sur un ou plusieurs buts à poursuivre.

2° Ces buts sont traduits en objectifs définis selon les grandes catégories comportementales, puis en objectifs opérationnels. De nouveau, l'accord doit se faire entre l'enseignant et l'inspecteur.

3° Un troisième accord préliminaire est conclu sur ce qui sera tenu pour preuve que l'apprentissage projeté a été réalisé par l'élève, qu'il s'agisse de connaissances, d'habiletés, d'attitudes.

Pareil « contrat », négocié pour des périodes variant entre une seule journée (voire une seule leçon) et plusieurs mois, donne à l'évaluation un rôle bien plus dynamique qu'une inspection de caractère impressionniste, souvent ressentie par l'enseignant comme une menace.

1° Ou bien le critère de réussite est précisé, ou bien les deux parties reconnaissent leur incapacité à évaluer objectivement ou, au moins, à se mettre d'accord sur les preuves de succès ou d'échec de l'enseignement.

2° L'inspecteur est tout de clarté, sur des observations faites directement sur la méthode de l'enseignant, sans critères d'évaluation qui, sans ambiguïté les limitent, attribuent à l'enseignant une responsabilité commune.

3° Les systèmes de formation des maîtres trouvent ici une utilisation nouvelle. Les traces du cheminement de l'enseignant sont prises en compte dans la stratégie en fonction de la situation, fonction d'*a priori* théorique.

Le système contractuel de formation des maîtres est un système qui a été pratiqué par des enseignants. Dans ce cas, ils « jouent » le jeu de l'évaluation mutuellement.

*La responsabilité civile contractuelle.* — Le système contractuel trouve une application nouvelle dans la notion de responsabilité (*accountability*) qui, depuis les années 1960, a été appliquée au milieu scolaire (Lessinger et Tyack, 1972).

La démarche fondamentale est :

1° Les objectifs de l'enseignant sont définis de façon opérationnelle.

2° On évalue dans ce contexte les élèves.

3° On rend compte de l'évaluation.

A partir de ce modèle, on peut définir dans certains systèmes de formation des maîtres :

(1) Pour l'analyse dans le contexte de la responsabilité, voir les meilleurs ouvrages existant sur ce sujet, notamment Rinehart & Winston, 1969.

(2) Les « contrats de performance » ont été utilisés dans le même contexte. Depuis 1969, dans le contexte de l'enseignement de certaines matières, comme les sciences ou mathématiques dans des zones défavorisées, le succès de l'enseignement, évalué en fonction de la réussite de faire et, en 1972, une enquête a révélé le caractère fallacieux, voire frauduleux, de ce système, cependant pas morte et, comme le système contractuel, la responsabilité civile a été ainsi plus facilement acceptée.

des expériences, que  
 ours de meilleurs résultats  
 es individus qu'il importe

fluence d'un enseignant  
 e ? D'une façon certaine,  
 t démontrer dans quelle  
 éral, et sa longévité sont  
 ir de cette constatation,  
 arquée au coin du bon  
 plines de base et de leur  
 pratique qui, d'ailleurs,  
 te. Pourquoi ne procé-  
 à partir de disciplines  
 age, les règles d'hygiène  
 mesurer correctement les  
 du tout ?

ouvement de recherche  
 des objectifs de l'éduca-  
 tives d'un grand intérêt.  
 iste déjà de nombreuses  
 Neil et Popham (1973,

at d'abord d'accord sur  
 éfinis selon les grandes  
 ctifs opérationnels. De  
 nant et l'inspecteur.  
 conclu sur ce qui sera  
 a été réalisé par l'élève,  
 l'attitudes.

diodes variant entre une  
 plusieurs mois, donne à  
 qu'une inspection de  
 par l'enseignant comme

écisé, ou bien les deux  
 luer objectivement ou,  
 es de succès ou d'échec

2° L'inspecteur est obligé de prendre *préalablement* position, en toute clarté, sur des objectifs qu'il préconise ou approuve, éventuellement sur la méthodologie qu'il recommande, et toujours sur les critères d'évaluation qu'il emploiera. Il est ainsi amené à reconnaître sans ambiguïté les limites de son savoir et donc à instaurer avec l'enseignant une relation coopérative, une action de recherche commune.

3° Les systèmes d'analyse des comportements d'enseignement trouvent ici une utilisation fonctionnelle : ils servent à garder certaines traces du cheminement didactique et à suggérer des variations de stratégie en fonction des performances des élèves (et non plus en fonction d'*a priori* théoriques ou d'idiosyncrasies inspectorales) (1).

Le système contractuel peut évidemment être appliqué lors de la formation des maîtres ou aussi en guise de système de perfectionnement pratiqué par des enseignants qui collaborent librement; dans ce cas, ils « jouent » le jeu des contrats et s'évaluent et se conseillent mutuellement.

*La responsabilité civile en matière d'enseignement.* — Le système contractuel trouve une application particulière — et vivement controversée — dans la notion de responsabilité civile en matière d'enseignement (*accountability*) qui, depuis 1970 surtout, apparaît dans la législation scolaire (Lessinger et Tyler, 1971; Gage, 1973) (2).

La démarche fondamentale est la suivante :

1° Les objectifs des programmes scolaires sont formulés de façon opérationnelle.

2° On évalue dans quelle mesure les objectifs sont atteints par les élèves.

3° On rend compte des résultats au public, aux autorités scolaires.

A partir de ce moment, des différences capitales interviennent. Dans certains systèmes scolaires, les résultats du bilan servent à

(1) Pour l'analyse dans le cadre de l'inspection, W. Popham estime que l'un des meilleurs ouvrages existant est : R. GOLDHAMMER, *Clinical supervision*, New York, Holt, Rinehart & Winston, 1969.

(2) Les « contrats de performance » (CAMPBELL et LORION, 1972) s'inscrivent dans le même contexte. Depuis 1969, des firmes commerciales américaines se chargent, par contrat, de l'enseignement de certaines branches à certaines populations d'élèves (par ex. lecture ou mathématiques dans des zones défavorisées). Le montant payé est proportionnel au succès de l'enseignement, évalué par des tests. On imagine les dangers de pareille façon de faire et, en 1972, une enquête officielle, portant sur dix-huit systèmes scolaires, a démontré le caractère fallacieux, voire frauduleux des « contrats de performance ». L'idée même n'est cependant pas morte et, comme le note GRONLUND (1974), le principe de la responsabilité civile a été ainsi plus facilement accepté.

stimuler des réformes, à justifier certains investissements nouveaux et, de façon générale, à prendre toutes mesures susceptibles de remédier aux faiblesses observées. Dans d'autres cas, le bilan devient celui du rendement de l'enseignant : s'il est très positif, une augmentation de traitement ou une promotion peut s'ensuivre; s'il est négatif, les mesures peuvent aller jusqu'au licenciement (voir, par exemple, le *Stull Act*, dans l'Etat de Californie) <sup>(1)</sup> <sup>(2)</sup>.

L'obligation de rendre publiquement compte de l'exécution suffisante des fonctions d'enseignement peut ne pas consister, ou ne pas uniquement consister en évaluation des apprentissages des élèves (*responsabilité des produits*) <sup>(3)</sup>; on distingue aussi la *responsabilité des processus* (dans quelle mesure l'enseignant connaît-il bien ses élèves et applique-t-il les méthodes et les matériels les plus susceptibles de produire de bons résultats?) et la *responsabilité expérimentale* (dans quelle mesure l'enseignant se renouvelle-t-il, introduit-il les apports de la recherche dans sa pratique?).

En général, les enseignants n'acceptent pas de bon gré d'être rendus *personnellement* responsables des résultats de leur activité; au-delà de nombreuses nuances, ils formulent cinq objections :

1° Si l'enseignant enseigne, il reste que l'élève doit apprendre et est donc aussi largement responsable (éventuellement avec sa famille) des progrès réalisés.

2° La mesure correcte et la juste appréciation des progrès réalisés sont très difficiles : manque de tests formatifs, difficulté d'évaluer dans

<sup>(1)</sup> *Stull Act (Assembly Bill No. 293 de l'Etat de Californie, 1971)* :

a) 13403. Aucun enseignant nommé définitivement ne peut être démis de ses fonctions, sauf pour une ou plusieurs des raisons suivantes : ... d) incompétence...

13487. Les autorités de chaque district scolaire élaboreront et adopteront des directives précises concernant l'évaluation; ces directives contiendront au moins :

- a) le niveau de progrès attendu chez les élèves dans chaque branche d'enseignement et les techniques permettant d'évaluer ce progrès;
- b) l'évaluation de la compétence de l'enseignant dans sa relation avec les standards établis;
- c) l'évaluation des autres prestations réclamées de l'enseignant comme appartenant à ses obligations normales;
- d) l'établissement des procédures et des techniques permettant de s'assurer que l'enseignant maintient une discipline adéquate et assure l'existence d'un environnement favorable à l'apprentissage.

13489. L'évaluation aura lieu au moins une fois tous les deux ans.

<sup>(2)</sup> Comme le note GRÖNLUND (1974), cette évaluation du rendement de l'école, dont l'IEA, puis le NAEP ont démontré la possibilité, est fortement approuvée par le public. Au sondage d'opinions Gallup de 1970, 67 % des adultes y étaient favorables aux Etats-Unis.

<sup>(3)</sup> Lier les augmentations de traitement, voire le traitement de base au rendement de l'enseignant, évalué par exemple par le nombre de calculs que les élèves pouvaient effectuer correctement en un temps fixé, a été pratiqué dès le XIX<sup>e</sup> siècle en Angleterre et aux Etats-Unis. Cette façon de faire a été combattue pour d'évidentes raisons.

les domaines du transfert (voir aussi, par exemple, l'ouvrage de GARDNER, 1975), importance de l'investissement <sup>(2)</sup>.

3° Le système encouragé par les apprentissages favorise des interactions plus affectives.

4° Le système crée la responsabilité avec les élèves.

5° Les pouvoirs organisés ont une responsabilité avec les enseignants, l'équipement, l'organisation, la recherche pédagogique et la formation.

Réagissant contre une évaluation purement pédagogique, certains demandent une responsabilité collective des enseignants devant les autorités scolaires, de la même façon que l'enseignant est responsable, à ce dernier niveau, de sa responsabilité directe et responsable, à l'égard de ses élèves.

Quelles que soient les raisons qu'elle soulève, la responsabilité médicale ne peut être évitée.

## I | L'ÉVALUATION PAR LES ENSEIGNANTS

Il est inutile d'insister sur le caractère purement subjective, auto-évaluative, contractuelle, recourant à l'auto-évaluation et de mesure, et suscitant une certaine résistance. De nombreux auteurs s'opposent à cette pratique.

<sup>(1)</sup> L'élève dont le résultat est inférieur à la moyenne, tandis que l'élève le meilleur tend à dépasser la moyenne, justifie les progrès, il importe de noter la moyenne.

<sup>(2)</sup> En certaines matières, les progrès peuvent être soit impossibles, soit très difficiles.

les domaines du transfert et de l'expression (V. et G. De Landsheere, 1975), importance de l'effet de régression (1) et de l'effet de plafonnement (2).

3° Le système encourage l'enseignant à concentrer ses efforts sur les apprentissages qui seront mesurés, au détriment d'autres interactions plus affectives.

4° Le système crée la tension et l'anxiété chez les enseignants.

5° Les pouvoirs organisateurs doivent porter conjointement la responsabilité avec les enseignants, car des pouvoirs dépendent l'équipement, l'organisation scolaire, l'organisation de centres de recherche pédagogique et de construction d'instruments d'évaluation.

Réagissant contre une conception étriquée de la responsabilité pédagogique, certains défendent le principe de la responsabilité collective des enseignants (De Landsheere, 1973) ou des enseignants, des autorités scolaires, des parents et des élèves (Havighurst, 1973). A ce dernier niveau, il serait plus exact de parler de participation, directe et responsable, à l'éducation scolaire.

Quelles que soient la difficulté et la complexité des problèmes qu'elle soulève, la responsabilité pédagogique (comme la responsabilité médicale) ne peut cependant pas être éludée.

## IV

### Qui évalue ?

#### I | L'ÉVALUATION PAR LES INSPECTEURS

Il est inutile d'insister sur la marge qui sépare une appréciation purement subjective, autoritairement formulée, d'une évaluation contractuelle, recourant à une batterie d'instruments d'observation et de mesure, et suscitant une réflexion et une recherche communes.

De nombreux auteurs s'accordent pour dénoncer les incohérences

(1) L'élève dont le résultat est le plus faible au prétest tend à faire le plus de progrès, tandis que l'élève le meilleur tend à faire peu de progrès, voire à régresser. Pour apprécier justement les progrès, il importe donc de tenir compte de l'effet de régression vers la moyenne.

(2) En certaines matières, les apprentissages au-delà d'un degré élevé de maîtrise sont, soit impossibles, soit très difficilement mesurables.

et les limites de validité de nombreuses évaluations inspectorales. McNeil et Popham (1973, p. 232) écrivent :

« Les effets de halo, le manque de définitions opérationnelles, l'absence d'échantillonnage rigoureux des comportements de l'enseignant, l'effet que la présence d'un inspecteur exerce sur le professeur sont autant de limitations qui jettent le doute sur les échelles d'évaluation maniées par les inspecteurs. »

Trois remarques s'imposent cependant :

a) Une expérience d'enseignement souvent longue, jointe aux apprentissages que les inspecteurs réalisent pendant l'exercice de leurs fonctions, en observant le travail d'un grand nombre d'éducateurs, permet généralement aux inspecteurs de détecter les meilleurs et les plus faibles des enseignants. Comme toujours, c'est l'évaluation fine des « moyens » qui fait problème.

b) Il serait injuste de reprocher aux inspecteurs de ne pas utiliser des méthodes objectives, alors que, nous venons de le voir, elles commencent seulement à exister.

c) Une observation approfondie, un diagnostic fin, une action de perfectionnement exigent un temps considérable. Jamais quelques inspecteurs ne pourront remplir ces tâches à l'échelle d'une nation : tout le système inspectoral doit être remis en cause, notamment à l'occasion de la mise en place de la formation continuée des enseignants.

## 2 | L'ÉVALUATION PAR LES COLLÈGUES

Il n'est pas rare que des enseignants portent un jugement sur la valeur professionnelle d'un collègue sans l'avoir systématiquement observé pendant son travail. Cette appréciation, souvent fondée sur des aspects marginaux, offre peu d'intérêt. Rosenshine (1971b) ne relève que deux recherches où l'appréciation par des collègues est mise en relation avec les performances des élèves. Les corrélations sont pratiquement sans signification.

Trop rarement, des enseignants collaborent pour s'évaluer mutuellement au point de vue du contenu enseigné ou de la méthodologie utilisée. Dans ce dernier cas, les divers instruments d'évaluation peuvent naturellement être employés et il semble que l'évaluateur bénéficie de l'observation autant que l'évalué. On manque cependant de résultats de recherches clairs à cet égard.

La fidélité (concordance) des évaluations portées par différents professeurs sur le contenu d'un même cours se révèle toutefois peu élevée. La raison fondamentale réside dans la difficulté de s'accorder

sur des critères, et la validité de ces critères en cause. En dernière analyse, il ne semble pouvoir fournir une mesure fiable. Et pour obtenir un coefficient de concordance en général, pas trop positif.

## 3 | L'ÉVALUATION PAR LES ÉLÈVES

Pratiquée depuis longtemps par leurs élèves, l'évaluation par les enseignants par leurs élèves dans nos pays, malgré des oppositions, semble être faite, soit pour des raisons d'efficacité, soit pour des raisons de simplicité donnant leur avis sur leur professeur de ce qu'ils ont appris. Il semble que pareille évaluation existe en France (Gage, 1972, p. 100).

Même injustifiée obligeamment, d'ailleurs une signification n'est pas un partisan de l'évaluation par les élèves sont hautement corrélés avec les performances parce que les valeurs de l'évaluation parce que l'opinion des élèves

### a | L'évaluation par les élèves

Essentiellement, l'évaluation par les élèves de la teneur du contenu enseigné, de la qualité psychologiques des enseignants, ne semble pas être en mesure de mesurer ces aspects, il en va de même.

En effet, on est en droit de se demander *hic et nunc* l'exactitude de l'évaluation du contenu avec l'ensemble des élèves n'aurait plus guère de rapport.

Certes, les élèves ont-ils appris ce qu'ils apprennent. Twyford (1971) a étudié dans une table d'une classe un dixième de la classe le taux d'apprentissage de choses exactes ? L'exactitude est, à cet égard, éclaircie par une conférence (intitulée « L'évaluation par les élèves »).

évaluations inspectorales.

initions opérationnelles, importements de l'ensei- exerce sur le professeur sur les échelles d'évalua-

ent longue, jointe aux pendant l'exercice de grand nombre d'éduca- de détecter les meilleurs ujours, c'est l'évaluation

ecteurs de ne pas utiliser enons de le voir, elles

nostic fin, une action de rable. Jamais quelques l'échelle d'une nation : en cause, notamment à ontinuée des enseignants.

tent un jugement sur la avoir systématiquement ion, souvent fondée sur Rosenshine (1971b) ne n par des collègues est élèves. Les corrélations

at pour s'évaluer mutuel- é ou de la méthodologie nstruments d'évaluation semble que l'évaluateur . On manque cependant

s portées par différents e se révèle toutefois peu a difficulté de s'accorder

sur des critères, et la validité des jugements est, par le fait même, aussi en cause. En dernière analyse, seule la concordance des jugements semble pouvoir fournir un indice immédiat de la validité du contenu. Et pour obtenir un coefficient de fidélité élevé, six collègues ne sont, en général, pas trop pour juger un même cours (Gage, 1974, p. 72).

### 3 | L'ÉVALUATION PAR LES ÉLÈVES

Pratiquée depuis longtemps aux Etats-Unis, l'évaluation des enseignants par leurs élèves devient de plus en plus fréquente dans nos pays, malgré des oppositions parfois farouches. L'évaluation peut être faite, soit pour des raisons administratives (étudiants des universités donnant leur avis avant la nomination d'un professeur), soit pour des raisons d'efficacité de l'enseignement (les élèves informent leur professeur de ce qu'ils pensent de son action pédagogique). Il semble que pareille évaluation soit possible dès la quatrième année primaire (Gage, 1972, p. 172).

Même injustifiée objectivement, une évaluation négative revêt d'ailleurs une signification psychologique. Gage (1974) écrit : « Je suis partisan de l'évaluation par les étudiants non pas parce qu'elles sont hautement corrélées avec le rendement des études, non pas parce que les valeurs des étudiants sont indiscutables, mais bien parce que l'opinion des étudiants est en elle-même importante. »

#### *a | L'évaluation par les élèves est-elle valide et fidèle ?*

Essentiellement, l'évaluation peut porter sur trois aspects : l'exactitude du contenu enseigné, les qualités pédagogiques et les qualités psychologiques des enseignants. Alors que l'élève ne semble généralement pas être en mesure de porter un jugement valide sur le premier de ces aspects, il en va tout autrement pour les deux autres.

En effet, on est en droit de penser que si l'étudiant pouvait apprécier *hic et nunc* l'exactitude de ce qui lui est enseigné et la relation de ce contenu avec l'ensemble des savoirs et des pouvoirs à conquérir, il n'aurait plus guère de raisons... de rester étudiant.

Certes, les élèves ont souvent une vive conscience des moments où ils apprennent. Twyford (1954) l'a démontré en installant sous chaque table d'une classe un dispositif permettant d'indiquer, à tout instant, le taux d'apprentissage estimé. Mais comment savoir si l'on apprend des choses exactes ? L'expérience de Naftulin, Ware et Donnelly (1973) est, à cet égard, éclairante. Ils ont préparé un acteur à faire une conférence (intitulée « L'application de la théorie mathématique des

jeux à l'éducation médicale »), conférence dans laquelle certaines notions exactes alterneraient avec des anecdotes humoristiques, des digressions sans rapport avec le sujet, des citations fantaisistes et des inexactitudes. Un questionnaire administré immédiatement après la conférence révèle des réactions en forte majorité positives. Aussi les chercheurs tirent la conclusion « que la satisfaction des étudiants, en matière d'apprentissages, ne représente peut-être guère que l'illusion d'avoir appris ». Ainsi se confirme le manque de validité quant au fond.

Par contre, il apparaît de plus en plus nettement que les élèves évaluent bien l'habileté didactique d'un enseignant et la qualité de la relation qu'il établit avec sa classe. Ces qualités, les étudiants ont l'occasion de les vérifier tous les jours. Ils savent s'ils reçoivent un enseignement clair ou confus, trop rapide ou trop lent, vif ou terne, richement illustré ou abstrait, etc. De même, ils ne se trompent pas sur l'accueil que le maître leur réserve, sur sa disponibilité, sa gentillesse, sa tolérance, sa compréhension, sa générosité, son équité.

De façon plus générale, Gage (1974) estime que les observations suivantes permettent d'espérer un jugement valide chez l'étudiant : faible corrélation (parfois nulle) entre, d'une part, l'appréciation portée par les élèves et, d'autre part, les résultats scolaires obtenus en fin de cycle, le sexe de l'enseignant, le nombre de ses publications, sa réputation de chercheur, le jugement de ses collègues, la difficulté du cours...

Quant à la fidélité, elle est souvent élevée pour les deux aspects où la validité des jugements semble établie. Elle est, par exemple, de .80 à .90 pour l'évaluation de la clarté des explications. Cette observation, déjà faite par Remmers en 1929, se répète depuis.

De travaux qui se poursuivent depuis 1927 sur la *Purdue Rating Scale for Instruction*, Remmers (1963) conclut notamment :

« — La moyenne des évaluations des enseignants par minimum 25 de leurs élèves est au moins aussi fidèle que les meilleurs tests d'aptitudes pédagogiques.

« — Dix ans après avoir terminé leurs études, les anciens étudiants accordent à dix caractéristiques d'enseignants une importance relative très proche de l'importance accordée lorsqu'ils étaient étudiants ( $\rho = .92$ ).

« — On observe un accord important ( $\rho = .40$  à  $.68$ ) entre les évaluations d'un même enseignant faites par des étudiants qui ont terminé leurs études depuis dix ans et les étudiants actuels.

« — La difficulté du cours n'influence pas l'évaluation par les élèves. »

b / Réalisation. — M  
différents selon que le  
des fins administratives  
diagnostiques.

Dans le premier cas  
suffire. Exemples (Smol)

1. Si vous souhaitiez suivre  
serait donné par le professeur

Influencerait très favora- blement votre décision	Influencerait votre décis- de façon a- favorable
--	---

2. Au total, le cours fait

Excellent — Très bon

Lorsque des étudiants  
et des faiblesses de son  
démarche beaucoup p  
de Purdue, de H. Remm  
d'une batterie compren

— dix échelles descrip

- intérêt manifesté
- attitude sympathique
- justice dans l'évaluation
- attitude libérale
- présentation de
- sens de la mesure
- confiance en soi
- maniérisme;
- présentation ou
- stimulation de la

— seize échelles numé

- adéquation de la
- liberté laissée da
- degré de satisfac
- immédiats);
- clarté des object
- étudiants;
- correspondance
- effectif;

... dans laquelle certaines anecdotes humoristiques, des citations fantaisistes et des réactions immédiates après la majorité positives. Aussi les insatisfactions des étudiants, en fait, n'ont guère que l'illusion d'une validité quant au

... nettement que les élèves de l'enseignant et la qualité de la formation, les étudiants ont conscience de ne pas savoir s'ils reçoivent un enseignement trop lent, vif ou terne, et, ils ne se trompent pas sur la disponibilité, sa générosité, son équité.

... l'opinion que les observations sont valides chez l'étudiant : d'une part, l'appréciation des résultats scolaires obtenus en fonction de ses publications, sa renommée auprès de ses collègues, la difficulté du

... vue pour les deux aspects de la question. Elle est, par exemple, l'objet de nombreuses explications. Cette question, se répète depuis 1927 sur la *Purdue Rating* et notamment :

... enseignants par minimum de la même façon que les meilleurs tests

... études, les anciens étudiants ont accordé une importance relative à la même chose qu'ils étaient étudiants

( $\rho = .40$  à  $.68$ ) entre les réponses données par des étudiants qui ont été des étudiants actuels.

... pas l'évaluation par les

b / Réalisation. — N. L. Gage (1974) recommande des contenus différents selon que les questionnaires d'évaluation sont destinés à des fins administratives (nomination, promotion...) ou à des fins diagnostiques.

Dans le premier cas, cinq ou six questions générales devraient suffire. Exemples (Smock et Cooks, 1973, cités par Gage) :

1. Si vous souhaitiez suivre un nouveau cours dans ce département, le fait que ce cours serait donné par le professeur X...

Influencerait très favora- blement votre décision	Influencerait votre décision de façon assez favorable	N'importerait pas	Influencerait votre décision de façon assez défavorable	Influencerait très défavora- blement votre décision
--	--	----------------------	--	--

2. Au total, le cours fait par X... est :

Excellent — Très bon — Bon — Satisfaisant — Pauvre — Très pauvre.

Lorsque des étudiants désirent informer leur professeur des forces et des faiblesses de son enseignement, telles qu'ils les perçoivent, une démarche beaucoup plus analytique s'impose. L'*Echelle d'évaluation de Purdue*, de H. Remmers (1960), répond à cette exigence. Il s'agit d'une batterie comprenant :

— dix échelles descriptives :

- intérêt manifesté par le professeur pour la matière enseignée;
- attitude sympathique envers les étudiants;
- justice dans l'évaluation des étudiants;
- attitude libérale;
- présentation de la matière;
- sens de la mesure et de l'humour;
- confiance en soi;
- maniérisme;
- présentation ou apparence;
- stimulation de la curiosité intellectuelle.

— seize échelles numériques :

- adéquation de la méthode d'enseignement;
- liberté laissée dans le choix de la matière;
- degré de satisfaction des besoins (objectifs ultimes et immédiats);
- clarté des objectifs du cours et discussion des objectifs avec les étudiants;
- correspondance entre les objectifs annoncés et l'enseignement effectif;



- adéquation du matériel didactique;
- adéquation de l'équipement des laboratoires;
- adéquation des notes de cours et des manuels;
- utilisation de tests formatifs;
- degré d'hétérogénéité des aptitudes des étudiants;
- adéquation de la quantité et du type de tâches à domicile;
- importance accordée aux tests pour décider du résultat final des élèves;
- correspondance entre les questions d'examens et les objectifs du cours;
- fréquence des tests;
- évaluation globale de l'enseignant.

On pourrait encore ajouter à cette liste :

- aptitude à répondre aux questions posées;
- façon de parler, prononciation;
- utilisation du tableau, du rétroprojecteur;
- aptitude à stimuler la participation des élèves;
- aptitude à stimuler l'intérêt des élèves.

*c | Effets du « feedback » provenant des élèves.* — Gage (1963) a construit un questionnaire à l'aide duquel des élèves de sixième primaire ont apprécié leur instituteur, puis précisé comment ils le souhaiteraient. Tous les mois, pendant une année scolaire, 86 instituteurs ont été informés de l'évaluation par leurs élèves; un groupe de contrôle de 90 instituteurs, bien qu'évalués de façon similaire, n'a pas reçu l'information. Sur 10 des 12 échelles, les élèves constatent que leur maître se rapproche de leur idéal; quatre des changements sont statistiquement significatifs par rapport au groupe de contrôle.

Neuf ans après cette expérience, Gage (1972), dressant le bilan de la recherche en la matière, estime cependant que, pour les enseignements primaire et secondaire, le *feedback* apporté par les élèves n'influence que modérément l'enseignant. La même remarque s'applique à l'enseignement supérieur. J. Centra (1972) est plus optimiste et conclut que l'influence de l'évaluation par les élèves peut déjà se manifester après trois mois. L'influence semble la plus nette lorsque les enseignants se sont auto-évalués beaucoup plus favorablement que leurs élèves.

Gage (1974) déconseille tout étalonnage normatif des évaluations et recommande vivement que l'appréciation d'une caractéristique par l'élève soit immédiatement suivie d'une invitation à préciser comment le professeur idéal devrait se comporter à ce propos.

## 4 | L'AUTO-ÉVALUATION

Tantôt introspection de systèmes de signes, tantôt objet de jugement est d'abord sujet des apprentissages, les élèves se font d'appréciation.

Howsam (1960) compare l'aide d'échelles d'appréciation.

Si des échelles permettent de constater la distance que l'enseignant se font d'eux-mêmes.

Plusieurs études rapportées par Howsam, initiées aux enseignants, pratiquent plus le style d'auto-évaluation après douze mois. Flanagan rapporte que les enseignants continuent à préférer au système d'auto-évaluation.

Il importe de la nuancer, car si l'enseignant éprouve le modèle proposé. Il semblerait, par exemple, augmentation de l'auto-évaluation impliquent une modification des habitudes pédagogiques; une guidance prolongée.

Les systèmes de contrôle à des fins d'auto-évaluation, de parler d'autoanalyse, de compte de ses propres erreurs, de devoir être rejetée aussi, s'ensuit visent uniquement des changements, de modèles personnels ou en fonction du modèle. Rosenshine et Flanagan rapportent que sans évaluation si le développement n'est pas mis en relation avec des apprentissages.

A notre avis, il est préférable d'un système d'analyse de se demander dans quelle mesure l'enseignant se demande dans le sens qu'il pense que l'enseignant se croie, d'en

## 4 | L'AUTO-ÉVALUATION

Tantôt introspection générale, tantôt observation de soi à l'aide de systèmes de signes ou de catégories, l'auto-évaluation de l'enseignant est d'abord subjective. Elle trouve une base plus objective quand les apprentissages, les performances des élèves deviennent le critère d'appréciation.

Howsam (1960) constate que les enseignants qui s'évaluent à l'aide d'échelles d'appréciation se surestiment généralement.

Si des échelles précises sont utilisées, les enseignants peuvent constater la distance qui sépare l'avis de leurs élèves de l'opinion qu'ils se font d'eux-mêmes.

Plusieurs études rapportées par Flanders (1969) montrent que les maîtres, initiés aux systèmes d'analyse d'interactions en classe, pratiquent plus le style indirect, le changement de style subsistant après douze mois. Flanders formule implicitement l'hypothèse que les enseignants continuent, dans leur pratique quotidienne, à se référer au système d'analyse. Bayer (1973) défend la même opinion. Il importe de la nuancer en ce sens que la modification n'intervient que si l'enseignant éprouve réellement le désir de se conformer au modèle proposé. Il semble aussi que certaines modifications (par exemple, augmentation du *feedback* spécifique apporté aux élèves) impliquent une modification tellement profonde des attitudes et des habitudes pédagogiques que l'enseignant ne peut y parvenir sans une guidance prolongée (Martin, 1970).

Les systèmes de catégories peuvent-ils être directement utilisés à des fins d'auto-évaluation ? Il serait probablement préférable de parler d'autoanalyse, c'est-à-dire d'un moyen de mieux se rendre compte de ses propres comportements. L'idée d'évaluation semble devoir être rejetée aussi longtemps que l'autoscopie et l'analyse qui s'ensuit visent uniquement à l'appropriation de modes de comportements, de modèles jugés désirables sur la foi d'une appréciation personnelle ou en fonction du prestige scientifique de l'auteur du modèle. Rosenshine et Furst (1973, pp. 151-152) estiment qu'il n'y a pas évaluation si le déploiement d'un comportement d'enseignement n'est pas mis en relation corrélationnelle ou, mieux encore, causale avec des apprentissages d'élèves.

A notre avis, il est cependant d'un haut intérêt que, disposant d'un système d'analyse opérationnel, un enseignant s'en serve pour se demander dans quelle mesure ses comportements sont bien orientés dans le sens qu'il pense leur donner. Il n'est, en effet, pas rare qu'un enseignant se croie, d'entière bonne foi, libéral, peu interventionniste,

plus centré sur l'élève que sur la matière... et fasse le contraire.

On ne semble pas avoir accordé assez d'attention à une observation faite par Rosenshine (1971b). Sur vingt et une recherches portant sur l'auto-évaluation, Rosenshine en trouve trois où les apprentissages des élèves sont pris comme critères et, dans les trois cas, les corrélations sont significatives, alors qu'elles ne le sont pas avec les appréciations par les étudiants, les directeurs d'école ou les inspecteurs. Rosenshine écrit (p. 179) :

« La cohérence des résultats, dans les recherches où les maîtres se sont auto-évalués, suggère des idées fascinantes pour les recherches futures. (...) Les résultats sont d'autant plus convaincants que les auto-évaluations ont été faites avec des instruments différents » (1).

### Conclusion

Il peut sembler que le gigantesque effort de recherche en matière d'évaluation de l'enseignement, déployé au cours des deux dernières décennies, débouche sur de maigres résultats. Comment en serait-il autrement dans un domaine aussi complexe ? Des solutions définitives, que seuls les rêveurs peuvent attendre, ne sont pas apparues, mais une importante clarification s'est opérée et des progrès modestes mais sûrs ont été réalisés.

Le mot enseignement recouvre de multiples phénomènes, résultant d'un grand nombre de facteurs objectifs et subjectifs, agissant isolément ou en interaction. Par conséquent :

1° Une évaluation correcte ne pourra vraisemblablement jamais se faire à l'aide d'un instrument unique et universel. C'est résolument vers une approche multidimensionnelle que l'on s'oriente.

2° On ne peut mesurer avec précision que des aspects particuliers. De vastes *constructs*, comme « enseignement frontal », « méthode traditionnelle », « enseignement par résolution de problèmes », etc., recouvrent tellement de choses différentes et se colorent de façons si diverses selon les individus et les situations qu'ils ne permettent aucune appréhension rigoureuse. L'échec presque intégral des recherches passées en témoigne.

(1) Mesure de la congruence entre le moi idéal et le *self concept* (Aspy); simple appréciation de sa propre efficacité en réponse à une seule question (Medley); simple appréciation de ses capacités en réponse à une seule question (McCall).

3° En raison de l'unicité de toute intention, jamais de lois universelles contemporaines sur les séquences d'actes pédagogiques possible : on observe les différences de rôles joués par les objectifs poursuivis (Gardner) ; un scientisme étriqué de l'enseignant continuera à être un facteur déterminant nouveau. L'imprévisibilité des comportements et de

- BAYER (E.), L'analyse des pratiques de l'enseignement, Paris, 1973, 30-40.
- BOND (G.) et DYKSTRA (R.), Instruction, in *Reading Research*, éd. par G. Bond et R. Dykstra, Buffalo, State Univ. of New York, 1972.
- CAMPBELL (R.) et LORION (J.), *Two Studies on the Effectiveness of Instruction*, ETS, 1972.
- DE BAL (R.), DE LANDSHEERE (G.), *Les tests de mesure descriptive*, Bruxelles, 1977.
- DELCHAMBRE (A.), *Mise en relation de la mesure et de la situation pédagogique*, Université de Bruxelles, 1977.
- DE LANDSHEERE (G.), *Introduction à la psychologie de l'enseignement*, 1976, 4<sup>e</sup> éd., 404 p.
- DE LANDSHEERE (G.), *Analysis of Teaching*, *Towards a Science of Teaching*, 1977.
- DE LANDSHEERE (G.), *Le test de mesure descriptive*, Bruxelles, novembre 1973.
- DE LANDSHEERE (G.), *Evaluation de l'enseignement*, Bruxelles, Labor, 1974, 30-40.
- DE LANDSHEERE (V. et G.), *Les tests de mesure descriptive*, Bruxelles, 1969 (3<sup>e</sup> éd., 1975).
- DETHIEUX (M.), et LECLERCQ (J.), *La mesure de l'efficacité de l'enseignement*, in *Recherches sur l'évaluation de l'Education nationale*, 1977.

... et fasse le contraire. L'attention à une observation et une recherche portant sur trois cas, les corrélations établies avec les appréciations des inspecteurs. Rosenshine a fait des recherches où les maîtres sont intéressés pour les recherches plus convaincantes que les résultats différents » (1).

3° En raison de l'énorme quantité de facteurs mis en jeu et de l'unicité de toute interaction humaine, il n'existera probablement jamais de lois universelles de l'enseignement. Beaucoup de recherches contemporaines sur l'évaluation tendent à se concentrer sur des séquences d'actes pédagogiques aussi bien définies et contrôlées que possible : on observe les effets sur les étudiants, en tenant compte des différences de rôles joués par les enseignants, du type d'élèves, des objectifs poursuivis (Gage, 1972, pp. 206-207). Bref, réagissant contre un scientisme étiqué qui a conduit dans une impasse, on pose que l'enseignant continuera à créer à chaque occasion un monde relationnel nouveau. L'important est donc de lui donner une palette de comportements et de moyens aussi riche que possible.

## BIBLIOGRAPHIE

de recherche en matière de cours des deux dernières années. Comment en serait-il ? Des solutions définitives, qui n'ont pas apparues, mais qui ont permis des progrès modestes mais constants. Ces phénomènes, résultant de processus subjectifs, agissant isolément et de façon inattendue, n'ont jamais été universels. C'est résolument vers l'avenir qu'on s'oriente. On se préoccupe de certains aspects particuliers. On parle de « méthode frontale », « méthode de problèmes », etc., qui se colorent de façons si diverses qu'ils ne permettent aucune synthèse intégrale des recherches effectuées. Le concept (Aspy) ; simple appréciation (Medley) ; simple appréciation (1).

- BAYER (E.), L'analyse des processus d'enseignement, in *Revue française de Pédagogie*, 24, 1973, 30-40.
- BOND (G.) et DYKSTRA (R.), The Cooperative research program in first-grade reading instruction, in *Reading Research Quarterly*, 2, 1967, 1-142.
- BURKHART (R. C.) (ed.), *The assessment revolution : New viewpoints for teacher evaluation*, Buffalo, State Univ. of New York, 1969.
- CAMPBELL (R.) et LORION (J.), *Performance Contracting in School Systems*, Columbus, Merrill, 1972.
- CENTRA (J.), *Two Studies on the Utility of Students Ratings for Instructional Improvement*, Princeton, ETS, 1972.
- DE BAL (R.), DE LANDSHEERE (G.), PAQUAY-BECKERS (J.), *Construire des échelles d'évaluation descriptives*, Bruxelles, Ministère de l'Éducation nationale, Organisation des Études, 1977.
- DELCHAMBRE (A.), *Mise en relation des comportements verbaux et non verbaux (visuels et auditifs) en situation pédagogique*, Université de Liège, 1975 (mémoire inédit).
- DE LANDSHEERE (G.), *Introduction à la recherche en éducation*, Paris, A. Colin; Liège, G. Thone, 1976, 4<sup>e</sup> éd., 404 p.
- DE LANDSHEERE (G.), Analysis of Verbal Interaction in the Classroom, in G. CHANAN (ed.), *Towards a Science of Teaching*, Londres, NFER, 1973.
- DE LANDSHEERE (G.), *Le test de closure*, Paris, Nathan; Bruxelles, Labor, 1973.
- DE LANDSHEERE (G.), Formes nouvelles de l'évaluation, in *Le Français dans le Monde*, octobre-novembre 1973.
- DE LANDSHEERE (G.), *Évaluation continue et examens. Précis de docimologie*, Paris, Nathan; Bruxelles, Labor, 1974, 3<sup>e</sup> éd.
- DE LANDSHEERE (V. et G.), *Définir les objectifs de l'éducation*, Paris, PUF; Liège, Thone, 1975.
- DE LANDSHEERE (G.) et BAYER (E.), *Comment les maîtres enseignent. Analyse des interactions verbales en classe*, Bruxelles, Ministère de l'Éducation nationale, Organisation des Études, 1969 (3<sup>e</sup> éd., 1975).
- DETHEUX (M.), et LECLERCQ-BOXUS (E.), Essai d'évaluation continue en éducation compensatoire..., in *Recherches sur les handicaps socio-culturels de 0 à 7-8 ans*, Bruxelles, Ministère de l'Éducation nationale, 1973.

- DUSSAULT (G.), LECLERC (A.-M.), BRUNELLE (J.), TURCOTTE (C.), *L'analyse de l'enseignement*, Montréal, Presses de l'Université du Québec, 1973.
- EBLE (K. E.), *The recognition and evaluation of teaching*, Salt Lake City, Univ. of Utah, 1970.
- EVANS (K.), Research on teaching ability, in *Educational Research*, I, 3, 1959.
- FAUQUET (M.) et STRASFOGEL (S.), *L'audio-visuel au service de la formation des enseignants*, Paris, Delagrave, 1972.
- FERRY (G.) et al., Les communications dans la classe, in *Bulletin de Psychologie*, XXII, 1-2, 1968, 81-95.
- FLANDERS (N. A.), Teacher Effectiveness, in *Encyclopedia of Educational Research*, New York, Macmillan, 1969.
- GAGE (N. L.), A method for improving teacher behavior, in *Journal of Teacher Education*, 14, 1963, 261-266.
- GAGE (N. L.), *Teacher effectiveness and teacher education*, Palo Alto, Pacific Books, 1972.
- GAGE (N. L.) (ed.), *Mandated Evaluation of Education*, Stanford, Stanford University, 1973.
- GAGE (N. L.), Ratings of college teaching, in N. S. GLASMAN et B. R. KILLAIT (ed.), *Second UCSB Conference on Effective Teaching*, Santa Barbara, Univ. of California, 1974.
- GALLAGHER (J.), *Teacher Variation in Concept Presentation in BSCS Curriculum Program*, Urbana, Univ. of Illinois, 1966.
- GALLAGHER (J.), Three studies of the classroom, in *Classroom Observation*, Chicago, Rand McNally, 1970.
- GRONLUND (N. E.), *Determining Accountability for Classroom Instruction*, New York, Macmillan, 1974.
- GUILFORD (J. P.), *Psychometric Methods*, New York, McGraw-Hill, 1954, 2<sup>e</sup> éd.
- HAVIGHURST (R.), Joint accountability, in A. ORNSTEIN (ed.), *Accountability for Teachers and School Administrators*, Belmont, Fearon, 1973.
- HENRY (G.), *Comment mesurer la lisibilité*, Paris, Nathan; Bruxelles, Labor, 1975.
- HILDEBRAND (M.) et WILSON (R.), *Effective University Teaching and its Evaluation*, Berkeley, Univ. of California, 1970.
- HILLER (J.), FISHER (G.) et KAESS (W.), A Computer Investigation of Verbal Characteristics of Effective Classroom Lecturing, in *American Educational Research Journal*, 6, 1969.
- HOWSAM (R.), *Who is a Good Teacher? Problems and Progress in Teacher Evaluation*, California Teachers Association, 1960. Cité par FLANDERS, 1969.
- JACKSON (P.), *The Way Teaching Is*, Washington, NEA, 1966.
- JOYCE (B.) et HAROO TUNIAN (B.), *The Structure of Teaching*, Chicago, SRA, 1967.
- LÉON (A.), Les grilles d'observation des situations pédagogiques, in *Revue française de Pédagogie*, 30, 1975, 5-13.
- LESSINGER (L.) et TYLER (R.) (ed.), *Accountability in Education*, Worthington, Jones, 1971.
- MARKLUND (S.) et GRAN (B.), La recherche et l'innovation en matière de formation des enseignants, in OCDE, *Tendances nouvelles de la formation et des tâches des enseignants*, Suède, 1974, 17-86.
- MARTIN (J.-M.), *Essai de modification contrôlée du comportement d'un instituteur. Etude des fonctions de feedback*, Université de Liège, 1970 (mémoire inédit).
- MCKEACHIE (W.), Research in teaching at college and university level, in N. GAGE, *Handbook of Research on Teaching*, Chicago, Rand McNally, 1963.
- MCNEIL (J.) et POPHAM, The assessment of teacher competence, in R. TRAVERS (ed.), *Second Handbook of Research on Teaching*, Chicago, Rand McNally, 1973.
- MIALARET (G.), La technique moderne et la formation des éducateurs, in *Les Sciences de l'Éducation*, 3, 1971, 4-48.
- MIALARET (G.), Analyse des situations éducatives, in M. DEBESSE et G. MIALARET, *Traité des sciences pédagogiques*, 4 : *Psychologie de l'éducation*, Paris, PUF, 1974.

- MITZEL (H. E.), *Teacher effectiveness*, New York, Macmillan, 1966.
- MORRISON (B.), *The Reaction of Teachers*, Univ. du Michigan, 1966.
- NAFTULIN (D.), WARE (J.) et WARE (J.), *Educational Seduction*, in N. GAGE, 1974.
- POSTIC (M.), L'analyse des acceptions de l'Éducation, 1, 1971, 57-66.
- POSTIC (M.), *Observation et formation*, in N. GAGE, 1974.
- POSTLETHWAITE (T. N.), *L'Évaluation de l'enseignement* (ronéotypé).
- REMMERS (H. H.), *Rating Method*, 1963.
- ROSENSHINE (B.), *New directions in teacher education*, Washington, USOE, 1971a.
- ROSENSHINE (B.), *Teaching Behavior*, in N. GAGE, 1974.
- ROSENSHINE (B.), *Objectively Measured Teacher Effectiveness*, Stanford, 1968 (thèse de doctorat).
- ROSENSHINE (B.) et FURST (D.), *Teacher Effectiveness*, in R. TRAVERS, *Second Handbook of Research on Teaching*, Chicago, Rand McNally, 1973.
- RYANS (D. G.), *Characteristics of Effective Teachers*, 1960.
- SIMON (A.) et BOYER (E.), *Ministry of Education*, Philadelphia, Research Center for the Study of Educational Administration, 1966.
- SPAIGHTS (E.), *Students' Appraisal of Teacher Effectiveness*, in N. GAGE, 1974.
- STAKE (R.), *New Trends in Evaluation*, in N. GAGE, 1974.
- TURNER, *Characteristics of Beginning Teachers*, in N. GAGE, 1974.
- TWYFORD (L. C.), *Profile of Teacher Effectiveness*, 2, 1954, 243-262.
- VERBIST (R.), et VAN HERREVEN, *Teacher Effectiveness*, in *Scientia Paedagogica*, 1966.

- E (C.), *L'analyse de l'enseignement*, Salt Lake City, Univ. of Utah, 1973.
- Research, I, 3, 1959.
- la formation des enseignants, Paris, PUF, 1974.
- ulletin de Psychologie, XXII, 1-2, 1974.
- Educational Research, New York, 1973.
- in *Journal of Teacher Education*, 1972.
- o Alto, Pacific Books, 1972.
- ford, Stanford University, 1973.
- N et B. R. KILLAIT (ed.), *Second Year Curriculum Program*, Urbana, Illinois, 1974.
- room Observation, Chicago, Rand McNally, 1973.
- struction, New York, Macmillan, 1960.
- aw-Hill, 1954, 2<sup>e</sup> éd.
- (ed.), *Accountability for Teachers*, Bruxelles, Labor, 1975.
- ing and its Evaluation, Berkeley, 1974.
- estigation of Verbal Characteristics, *Educational Research Journal*, 6, 1974.
- in *Teacher Evaluation*, California State University, 1967.
- gogiques, in *Revue française de Pédagogie*, Worthington, Jones, 1971.
- on en matière de formation des enseignants, *Etude des tâches des enseignants*, Suède, 1974.
- d'un instituteur. *Etude des fonctions*, in N. GAGE, *Handbook of Teacher Education*, University level, in N. GAGE, *Handbook of Teacher Education*, 1963.
- petence, in R. TRAVERS (ed.), *Second Year Curriculum Program*, Rand McNally, 1973.
- es éducateurs, in *Les Sciences de l'Éducation*, Paris, PUF, 1974.
- MITZEL (H. E.), Teacher effectiveness, in C. W. HARRIS, *Encyclopedia of Educational Research*, New York, Macmillan, 1960.
- MORRISON (B.), *The Reaction of Internal and External Children to Patterns of Teaching Behavior*, Univ. du Michigan, 1966 (thèse de doctorat). Cité par FLANDERS, 1969.
- NAFTULIN (D.), WARE (J.) et CONNELLY (F.), The Doctor Fox Lecture : A Paradigm of Educational Seduction, in *Journal of Medical Education*, 48, 1973, 630-635. Cité par GAGE, 1974.
- POSTIG (M.), L'analyse des actes pédagogiques des professeurs de sciences, in *Les Sciences de l'Éducation*, 1, 1971, 57-119.
- POSTIG (M.), *Observation et formation des enseignants*, Paris, PUF, 1977.
- POSTLETHWAITE (T. N.), *L'éducation des maîtres et l'efficacité des maîtres*, Paris, IJPE, 1974 (ronéotypé).
- REMMERS (H. H.), Rating Methods in Research on Teaching, in N. L. GAGE, *Handbook...*, 1963.
- ROSENSHINE (B.), New directions for research on teaching, in *How Teachers make a Difference*, Washington, USOE, 1971a.
- ROSENSHINE (B.), *Teaching Behaviours and Student Achievement*, Londres, NFER, 1971b.
- ROSENSHINE (B.), *Objectively Measured Behavioral Predictors of Effectiveness in Explaining*, Stanford, 1968 (thèse de doctorat). Cité par GAGE (1972), p. 122.
- ROSENSHINE (B.) et FURST (N.), The use of direct observation to study teaching, in R. TRAVERS, *Second Handbook of Research on Teaching*, Chicago, Rand McNally, 1973.
- RYANS (D. G.), *Characteristics of Teachers*, Washington, American Council on Education, 1960.
- SIMON (A.) et BOYER (E.), *Mirrors for Behavior : An Anthology of Classroom Observation Instruments*, Philadelphia, Research for Better Schools, 1967 et 1970.
- SPAIGHTS (E.), Students appraise teachers' methods and attitudes, in *Improving College and University Teaching*, 15, 1967, 15-17.
- STAKE (R.), *New Trends in Evaluation*, Univ. of Göteborg, School of Education, 1974.
- TURNER, Characteristics of Beginning Teachers, in *School Review*, 73, 1965, 48-58.
- TWYFORD (L. C.), Profile Techniques for Program Analysis, in *Audiovisual Communication*, 2, 1954, 243-262.
- VERBIST (R.), et VAN HERREWEGHE (M. L.), Evaluation du comportement dans l'enseignement, in *Scientia Paedagogica Experimentalis*, XI, 2, 1974, 228-258.