

Systèmes de normes et évaluation scolaire dans le cycle fondamental

M. Derghal

Professeur à l'IPSE - Université d'Alger

I ntroduction

EN milieu scolaire, la notation est un phénomène psychologique central. La " note » constitue, pour le maître, une intention de sanction de l'élève, pour une activité déterminée. Elle représente à la fois un jugement pour le maître et une information pour l'élève. Elle est d'autant plus importante que l'on se trouve dans des systèmes d'éducation et d'enseignement fondés essentiellement sur la technique de « dressage », c'est-à-dire des formes particulières de systèmes de sanctions extrinsèques. (Derghal, 1985, 1992). Faut-il rappeler ici les travaux sur la docimologie, notamment ceux de H. Pieron (1963), et ceux de G. Noizet (1961, 1969, 1978), lesquels mettent l'accent sur l'importance du rôle joué par les facteurs affectifs associés aux facteurs cognitifs. H. Pieron (1963) parle d'un effet de halo qui se traduit généralement par une surestimation des « bons élèves » et une sous-estimation des « mauvais élèves ». On peut alors imaginer les conséquences et les effets psychologiques sur les comportements de ces élèves. A ce propos, on pourra se reporter à « l'effet Pygmalion » décrit par R. A. Rosenthal et L. Jacobson (1971) qui ont montré comment l'attente d'une personne à l'égard du comportement d'une autre personne peut, tout à fait inconsciemment se révéler exacte du simple fait qu'elle existe.

Pour les auteurs, il existe un rapport direct entre le succès des élèves et le comportement psychologique des enseignants vis-à-vis de cet éventuel succès.

Ils ont réalisé une expérience dans une école publique élémentaire d'un quartier pauvre de San Francisco.

20 % des élèves de l'établissement furent présentés à leurs maîtres comme capables d'un développement particulièrement brillant. Les noms de ces élèves ont été tirés au hasard.

Huit mois plus tard, on constatait chez ces enfants « miracles » un démarrage scolaire réel que confirmait une augmentation très significative de leur Q. I. par rapport à leurs camarades non signalés à l'attention de leurs maîtres. Le changement très favorable, d'attitude des maîtres à propos des performances intellectuelles de ces prétendus « enfants spéciaux » choisis au hasard, avait provoqué chez ces derniers des progrès intellectuels réels.

Pour G. Noizet (1961), les différences entre les correcteurs ne sont pas des différences fortuites, mais des différences de conceptions sur les buts mêmes de l'examen et les principes de notation.

G. Noizet et J. J. Bonniol (1969) considèrent l'évaluation scolaire comme un comportement, au sens que l'on donne en psychologie à ce mot, c'est-à-dire une manière globale de répondre à une situation dans laquelle on se trouve.

G. Noizet et J. P. Caverni (1978) ont procédé à une analyse expérimentale du comportement d'évaluation ; cette analyse consistait à identifier les variables de situation et à en mesurer l'importance, l'objectif étant d'aboutir à une explication causale de l'évaluation.

Les auteurs présentent l'évaluation comme une activité de comparaison entre une production scolaire à évaluer et un modèle de référence. Cependant, le modèle de référence est susceptible de varier, d'une part selon les copies qui dans le lot, précèdent la copie qui va être évaluée (c'est ce que Noizet appelle « effets d'ancrage »), et d'autre part, selon les informations dont l'évaluateur dispose à propos de l'auteur de la production qu'il va noter.

La signification de la note fait en réalité l'objet d'un débat chronique, implicite ou explicite, entre les enseignants et les élèves. Ces derniers contestent toujours le principe de « justice » et « légitimité » de la note.

Il est intéressant de signaler à ce propos une conception des sanctions didactiques qui a cours auprès de certains maîtres et que nous avons pu observer. Ceux-ci distribuent des « mauvaises » notes au début de l'année dans le but de « fouetter », « d'activer » les élèves, mais avec le résultat de pouvoir s'attribuer

le mérite des « progrès » de la classe en cours d'année, et d'améliorer ainsi, consciemment ou non, la sanction publique de leur propre comportement professionnel.

1 - EXPERIENCE

Ce n'est pas tout à fait la docimologie classique qui nous intéresse, mais la façon de noter du maître, et sa relation avec celle des élèves, dans la mesure où cette relation peut jouer un rôle important en matière d'imitation didactique ou de démonstration pédagogique.

Il est clair que la relation entre le système d'appréciation d'un maître et celui de ses élèves varie en fonction même de l'activité considérée. Ainsi, nous pouvons faire l'hypothèse facile que le sens de cette relation sera différent selon qu'il s'agira de mathématique ou bien de littérature. En considérant le niveau scolaire auquel appartiennent nos sujets d'expérience, nous parlerons de « problème » et de « rédaction ».

En effet, évaluer un problème, c'est évaluer le résultat d'un apprentissage particulier. Il existe, dans ce cas, une seule réponse attendue, qui peut-être donnée par certains, ou par tous les élèves, selon, bien sûr l'efficacité de l'enseignement reçu en cette matière. Pour chaque énoncé de problème, il existe une référence théorique précise qui détermine la réponse. L'appréciation, la notation, consistera alors, à comparer cette réponse attendue avec celles fournies par les élèves et à « mesurer » la distance qui existe entre ces deux types de réponses, suivant un barème normatif mais définissable. On peut donc faire l'hypothèse que, dans ce domaine, le système d'appréciation du maître et celui des élèves seront plus congruent, plus convergents que dans les domaines à faible univocité normative.

Par contre, lorsqu'on évalue une rédaction, on évalue surtout un comportement verbal, dans lequel un facteur subjectif diversifié est supposé intervenir. Pour un même thème proposé, il n'existe pas de réponse précise attendue, comme pour le problème. On admet que deux élèves, ou même deux maîtres, ne produiront jamais deux réponses identiques. Il s'ensuit qu'il existe autant de réponses attendues que de maîtres, et même plusieurs réponses attendues pour un même maître. La distance d'avec les réponses observées chez les élèves est difficile à évaluer ; c'est ce qui rend complexe l'appréciation d'une rédaction.

Il y aura possibilité de divergence entre les deux systèmes d'appréciation considérés (celui du maître et celui des élèves).

L'expérience réalisée comporte deux phases, lesquelles, quoique faisant appel à des méthodes différentes, n'en sont pas moins complémentaires, et ont été menées simultanément.

La première phase, au moyen d'une méthode « directe » (expérimentale) consistait à faire apprécier par les sujets leurs propres devoirs (auto-évaluation) ainsi que des « devoirs-types » composés par l'expérimentateur (hétéro-évaluation). Pour chacune des deux activités scolaires considérées, deux devoirs-types ont été élaborés à partir d'une sélection opérée parmi l'ensemble des devoirs produits par les sujets. Le premier devoir, que nous avons appelé devoir « Fond » réunit les critères de fond (critères opérationnels pour la tâche) et ne tient pas compte des critères de forme. Il est utile de rappeler que ce type de devoir est généralement considéré comme étant le « bon » devoir. Le deuxième devoir, que nous avons appelé devoir « Forme », par opposition au devoir fond, néglige les critères de fond, mais respecte par contre les critères purement formels (comme le style, la présentation, l'écriture, etc.) Ce type de devoir correspond plutôt à ce qu'on considère généralement comme un « mauvais devoir ».

La deuxième phase, (« indirecte ») consiste en une sorte de questionnaire post-expérimental, destiné à examiner et à analyser la manière d'évaluer les élèves ainsi que l'importance qu'ils ont accordée aux différents critères selon l'ordre d'importance qu'il leur accordait.

1. 1. HYPOTHESES

H1) La catégorisation (critères de fond, critères de forme) peut être tout à fait pertinent pour des activités comme le problème de Maths ; les sujets auront tendance à accorder plus d'importance aux critères de fond.

H2) Cette catégorisation ne sera pas très pertinente pour des activités comme la rédaction ; les sujets auront tendance à accorder autant d'importance aux deux types de critères qu'ils n'arriveront pas à départager.

H3) Pour le problème, les sujets attribueront des notes plus élevées au devoir « Fond » qu'au devoir « Forme ».

H4) Pour la rédaction, ils attribueront des notes à peu près égales aux deux devoirs-types.

H5) Pour le problème, il y aura homogénéité entre le système d'appréciation du maître et celui des élèves.

H6) Pour la rédaction, par contre, il y aura « hétérogénéité » entre ces deux systèmes d'appréciation.

1.2 PLAN D'EXPÉRIENCE ET DISPOSITIF EXPÉRIMENTAL

Notre expérience a porté sur 56 sujets du sexe masculin tirés au hasard sur 87 sujets de deux classes de 6ème année du cycle fondamental dans une école du centre d'Alger.

Les variables indépendantes :

1- « Activité scolaire » : 2 modalités ; problème, rédaction.

2- « Évaluation » (appréciation) ; 2 modalités ; auto-évaluation, hétéro-évaluation.

3- « Types de critères de construction de l'objet évalué » ; 2 modalités ; fond, forme.

La variable dépendante : La note chiffrée (de 0 à 10).

Pour contrôler les effets « d'ordre de passation » dus notamment aux variables : « activité scolaire » et « Type d'évaluation », nous avons constitué quatre situations expérimentales de 14 sujets chacune, et dont l'ordre de déroulement des différentes phases de l'expérience, s'établit comme suit :

Situation 1 :

Problème : Auto-évaluation, Hétéro-évaluation, Questionnaire.

Rédaction :

Situation 2 :

Rédaction :

Problème :

Situation 3 :

Problème : Hétéro-évaluation, Auto-évaluation, Questionnaire

Rédaction :

Situation 4 :

Rédaction :

Problème :

1.3. Déroulement de l'expérience

Deux jours avant la séance d'expérimentation proprement dite, nous avons proposé aux sujets, par l'intermédiaire des maîtres, un énoncé de problème et un thème de rédaction.

Nous avons ensuite demandé aux maîtres de noter les devoirs, mais en les priant toutefois de ne pas porter d'annotation sur les feuilles des élèves.

Présentation de l'expérience

« Aujourd'hui, c'est vous qui allez noter vos devoirs : le problème et la rédaction que vous avez faits il y a deux jours. Vous noterez aussi deux autres problèmes et deux autres rédactions de deux élèves qui ont fait le même travail que vous. Vous donnerez une note de 0 à 10, vous l'écrirez dans la marge et vous indiquerez pourquoi vous donnez cette note ».

L'expérimentateur s'assure de la compréhension de la sollicitation verbale par tous les sujets ; ensuite il distribue les premiers devoirs à évaluer. Lorsque tous les sujets ont noté ces premiers devoirs, l'expérimentateur les récupère, puis distribue les deuxième devoirs, et ainsi de suite.

Rappelons que chaque sujet avait trois devoirs à noter dans chacune des deux activités scolaires concernées : son propre devoir et les deux « devoirs-types » (devoir « Fond », devoir « Forme »).

Lorsque les sujets ont noté la première série de trois devoirs dans l'une des deux activités l'expérimentateur leur présente le questionnaire correspondant. Les sujets doivent alors attribuer un rang à chacun des critères en fonction de l'importance qu'ils leur accordent.

Après cela, on procède de manière identique pour l'autre activité.

2. Résultat et Analyse

2. 1. Hétéro-évaluation : Évaluation des devoirs-types»

Il s'agissait, pour les sujets, d'attribuer une note de 0 à 10 pour chacun de ces devoirs. Pour chaque activité, il y avait deux devoirs-types : un devoir « Fond » (bon devoir) et un devoir « Forme » (mauvais devoir).

Le tableau 1 rassemble les moyennes des notes attribuées par les sujets.

TABLEAU 1

(Moyennes des notes attribuées par les sujets aux « devoirs-types » : hétéro-évaluation)

Activités Scolaires SITUATIONS	PROBLEME		REDACTION	
	Devoir fond (Bon devoir)	Devoir forme (Mauvais devoir)	Devoir fond (Bon devoir)	Devoir forme (Mauvais devoir)
SITUATION 1 (A.P) Auto-évaluation d'abord Problème d'abord	9,42	3,42	7,89	7,14
SITUATION 1 (A.R) Auto-évaluation d'abord Rédaction d'abord	8,96	1,46	7,53	7,64
SITUATION 3 (H.P) Hétéro-évaluation d'abord Problème d'abord	8,03	4,04	5,52	5,21
SITUATION 4 (H.R) Hétéro-évaluation d'abord Rédaction d'abord	7,96	1,03	5,78	6,17
MOYENNE PAR DEVOIR	8,59	2,50	6,66	6,54

Sources de variation	Degrés de Liberté	Somme des carrés	Carrés Moyens	F de Snédécór	
INTER-GROUPES (G)	3	98,50	32,83	5,95	*
ordre					
Auto-évaluation/ hétéro-évaluation	1	83,79	83,79	15,20	**
ordre					
Problème/Rédaction.	1	14,50	14,50	2,63	—
Interaction	1	0,21	0,21	0,03	—
résidu S(G)	52	286,66	5,51		
INTER-DEVOIRS (T)	3	1.103,80	377,93	85,53	**
Problème/Rédaction.	1	62,16	62,16	14,45	**
Problème (Fond/Forme)	1	1.041,27	1.041,27	242,01	**
rédaction (Fond/Forme)	1	0,37	0,37	0,08	—
Interaction	9	127,19	14,13	3,28	**
Résidu S(G)	156	671,06	4,30		
TOTAL	223	2.287,21			

— : Non significatif

* : Significatif au seuil de $P < 0,005$

** : Significatif au seuil de $P < 0,001$

Les résultats montrent qu'ils ont opéré une différenciation dans la notation des deux activités scolaires.

En effet, pour le problème nous relevons une très nette différence entre les moyennes des notes attribuées au «bon devoir» et celles des notes attribuées au «mauvais devoir». La moyenne des notes du premier devoir est plus élevée que celle des notes du second devoir, et ceci dans les quatre situations expérimentales.

Par contre, pour ce qui est de la rédaction, les moyennes sont très rapprochées ; les sujets des situations 2 et 4 ont même accordé, en moyenne, des notes un peu plus élevées au «mauvais devoir» qu'au «bon devoir». Ainsi, il semble en ce qui concerne cette matière, que les sujets n'arrivent pas à distinguer, à différencier les deux types de critères.

Avant de pousser plus loin l'analyse et l'interprétation de ces résultats, il est indispensable de les soumettre à un test statistique et adéquat : l'analyse de variance (cf. tableau 2).

L'analyse de variance laisse apparaître, en premier lieu, un effet de la variable, «ordre de déroulement» des différentes phases de l'expérience, c'est à dire : problème, puis rédaction et réciproquement ; auto-évaluation, hétéro-évaluation et réciproquement.

($F = 5,95$ - significatif au seuil $P < 0,005$).

Cependant, dans le but de préciser davantage cet effet, nous avons décomposé le facteur en trois éléments.

Nous constatons alors que cet effet est dû essentiellement au type d'évaluation, c'est-à-dire à la différence : auto-évaluation, hétéro-évaluation.

Le fait de commencer par l'évaluation de sa propre activité joue un rôle dans l'appréciation ultérieure des autres individus. Il altère l'échelle de valeurs ou barème. Les résultats obtenus sur ce point font apparaître un phénomène qui mérite d'être relevé. Les sujets qui ont commencé par noter leur propre devoir (auto-évaluation - situation -1 et 2) ont attribué «aux autres» en général, et en moyenne, des notes plus élevées que ceux qui ont commencé par noter les devoirs-types (supposés provenir d'autres sujets).

Situation 1 : $m = 6,97$

Situation 2 : $m = 6,40$

Situation 1 et 2 : $m = 6,68$.

Situation 3 : $m = 5,68$

Situation 4 : $m = 5,24$

Situations 3 et 4 : $m = 5,46$.

Ces résultats montrent que dans une activité d'évaluation, des élèves (âgés de 11 à 13 ans) se comportent différemment selon qu'ils sont ou non impliqués dans le processus. Lorsqu'ils sont eux même concernés ou visés (cibles), par le système d'évaluation qu'ils utilisent, ils manifestent une certaine propension à la «générosité» (auto-indulgence; endonomie) et accordent en général des notes plus élevées. Le barème est plus élevé ou moins exigeant s'ils l'ont construit à leur propre usage.

L'analyse de variance montre également un autre effet celui de la variable «type de devoirs» (problème-rédaction, Fond/Forme) ; ($F= 85,53$. significatif au seuil de $P < 0,001$).

Si nous décomposons ce facteur en trois éléments, nous remarquons que cet effet est dû :

Critères Rang	a Thème	b Idées	c Correction	d Vocabulaire	e Longueur	f Présentation
Rang 1	17	15	16	8	6	11
Rang 2	10	15	14	11	11	6
Rang 3	12	15	14	10	9	6
Rang 4	4	8	9	18	7	12
Rang 5	8	4	7	9	9	18
Rang 6	10	4	1	5	19	8

*** TABLEAU 3**

(Hiérarchisation des critères de notation de la rédaction - Effectifs -)

1°) à la différenciation opérée par les sujets entre la rédaction et le problème ; (F=14,45. significatif au seuil de $P < 0,001$).

2°) à la distinction Fond/Forme, c'est-à-dire « bon devoir / mauvais devoir » ; mais cela seulement pour le problème ; (F=242,01. significatif au seuil de $P < 0,001$). d'où l'interaction. La rédaction ne donne lieu à aucune distinction due aux critères Fond/Forme.

Ainsi, les résultats obtenus dans ce domaine confirment nos hypothèses 1, 2, 3, 4 (cf. 1. 1).

2.2. Dépouillement des questionnaires

Il s'agissait de classer les critères de construction de l'objet évalué selon leur ordre d'importance.

Questionnaire sur la rédaction : les critères, étaient explicités de la manière suivante :

- Critères de Fond :** a : respecter le thème proposé, ne pas sortir du sujet ;
b : avoir des idées sur le thème ;
c : écrire des phrases correctes.
- Critères de Forme :** d : vocabulaire utilisé (riche, élaboré, ou pauvre) ;
e : dimension de la rédaction (longue, courte) ;
f : présentation, écriture.

Questionnaire sur le problème :

- Critères de Fond :** a : trouver les réponses justes
b : bien réfléchir, bien raisonner.
c : rédiger correctement les solutions.
- Critères de forme :** d : écrire l'unité qui convient à chaque nombre ;
e : présentation

* Le tableau 3 présente les résultats concernant le questionnaire sur la rédaction.

Si nous considérons les trois premiers rangs pour chacun des critères, nous constatons, pour ce qui est :

1°) des critères de fond: critères a : 39 sujets l'ont classé dans les rangs 1, 2 ou 3

" b : 45 " " " " "

" c : 44 " " " " "

Ainsi, 63% à 73% des sujets ont classé les critères de fond dans les trois premiers rangs.

2°) des critères de forme :

Critères d : 29 sujets l'ont classé dans les rangs 1, 2 ou 3.

e : 26

f : 23

37 à 47% des sujets ont classé les critères de forme dans les trois premiers rangs.

Ces résultats signifient que les sujets ont bien fait une distinction entre les différents types de critères de construction de l'objet évalué, et accordé une plus grande importance aux critères de fond.

Il semblerait alors qu'il y ait une contradiction entre ces résultats et ceux obtenus dans l'évaluation des devoirs-types, où les sujets ont accordé des notes sensiblement identiques aux devoirs (fond et forme). On pourrait sans doute l'expliquer par le facteur déjà signalé plus haut : l'évaluation, la notation d'une activité comme la rédaction est rendue plus difficile par l'absence de référence précise, «unique».

Les enfants contrairement à l'intention de l'adulte et son système d'appréciation sont attirés par la «surface» de la rédaction et les critères «superficiels» sur lesquels l'accord est facile : présentation, vocabulaire, etc. ; où il y a une attente de similitude. A priori, ils reconnaissent la prépondérance du fond, mais dans la pratique, ils se comportent différemment. En fait, ils sont divisés entre deux échelles de valeurs dont l'une (formelle) tend à primer plus dans la «pratique» à cause de son automatisme et de sa facilité technique, tandis que la qualité de fond peut-être accidentelle, liée au thème.

Nous rappelons que le thème de la rédaction permettait de faire intervenir la notion de valeurs morales ; les différents comportements relatés dans les rédactions étaient appréciés en fonction de ces valeurs. De plus, afin de permettre l'explication par les sujets des phénomènes, nous leur avons demandé de justifier leur note par un commentaire écrit. Les différentes observations rapportées par les sujets sont très édifiantes à cet égard.

TABLEAU 4
(Hiérarchisation des critères de notation du problème - Effectifs)

Critères Rang	a Réponses justes	b Raisonnement	c Correction	d Unité adéquante	e Présentation
Rang 1	19	15	10	13	10
Rang 2	5	22	11	12	11
Rang 3	12	11	15	14	9
Rang 4	13	8	11	13	13
Rang 5	12	5	14	9	18

Ainsi pour la justification des notes élevées, nous relevons des remarques telles que : « c'est un gentil garçon » ; « il aime les animaux » ; « il a bon coeur » ; « il a fait ce qu'il fallait ». Et pour justifier de mauvaises notes : « ce garçon n'a pas bon coeur » ; « il est dur » ; « il n'aime pas les animaux » ; « il n'a pas bien fait ».

L'expérience ne permet pas de traiter ce point à fond car la « moralité » du thème n'était pas une variable, mais une donnée stable. D'autres expériences pourraient l'utiliser comme variable. Cependant, là encore, on peut penser que l'effet de halo « moral » est plus fort chez l'enfant que chez l'adulte et constitue entre eux une divergence de normes. Les enseignants admettent en général que le type de justification morale est donné par les enfants n'est pas pertinent. Les enfants tombent facilement dans le piège du « moralisme » tendu par l'expérimentateur.

Le tableau 4 présente les résultats relatifs au dépouillement du questionnaire concernant le problème. Si nous considérons toujours les trois premiers rangs pour chacun des critères, nous remarquons :

1°) Critères de fond :

Critère a : 36 sujets l'ont classé dans les rangs 1, 2 ou 3.

b : 48

c : 36

2°) Critères de forme :

Critère d : 39 sujets l'ont classé dans les rangs 1, 2 ou 3.

e : 30

Contrairement à ce qui c'est passé pour la rédaction, les sujets ont accordé sensiblement la même importance subjective et explicite aux deux types de critères. Selon nous, cet état de fait tient à ce que, pour le problème, il n'y a pratiquement pas de critères de «forme» ; l'important étant de trouver la solution ; le résultat est «juste» ou «faux». Ainsi, peut-on expliquer la relative aisance avec laquelle ont été notés les devoirs-types dans ce domaine.

Les résultats obtenus sont très caractéristiques à cet égard.

Une telle activité repose essentiellement sur des éléments «opérationnels» qui font que l'on peut avoir une réponse précise attendue, qui servira de référence, d'«étalon» en quelque sorte, pour l'application d'un système d'évaluation. IL y a là une possibilité d'homogénéité entre des systèmes différents (par exemple, entre celui du maître et celui des élèves).

2.3 Comparaison des notes du maître avec celles des élèves.

Rédaction :

Notes attribuées par le maître : $m = 5, 31$.

les élèves : $m = 6, 33$

Comparaison des moyennes : t (Student - Fisher) = 2, 97.

Significatif au seuil de $P < 0,001$.

Problème :

Notes attribuées par le maître : $m = 7, 20$.

les élèves : $m = 7, 24$.

Il va sans dire que la comparaison des deux moyennes par le test du t (Student - Fisher) n'est pas significative. ($t = 0, 08$).

3 - Discussion

Cette expérience nous a permis de voir apparaître, dans un domaine particulier (notamment le domaine éthique) une divergence de normes, entre le maître et les élèves. Cette divergence a engendré des systèmes d'appréciation,

et donc de sanctions, différents. En effet, si entre les élèves, il y a une plus grande dispersion, par contre, ce qui est plus important, il y a une divergence d'appréciations entre le maître qui s'est montré plus « sévère » et les enfants qui se sont montrés plus généreux dans leur notation. De plus, le maître n'aurait probablement jamais eu une éthique de la conduite à l'égard du chien. Or, les enfants l'ont interprétée comme une question d'éthique (ce qui est bien, ce qui est mal, le bon garçon, le mauvais garçon, etc).

Lorsque cette divergence de normes est ressentie par les enfants, lorsqu'ils en recueillent les indices, elle peut créer chez eux une réaction d'« anti-apprentissage », une réaction de contre-imitation, une divergence systématique entre eux et le maître.

Il serait utile d'étudier comment se manifeste cette réaction de contre-imitation chez les enfants. Pour notre part ; nous proposons l'hypothèse qui consiste à penser que c'est en « créant » que peuvent se manifester des réactions collectives d'opposition, de la part des élèves.

Ceci dans la mesure où ils auront le sentiment que, à travers les activités du maître. (notamment les activités des notations), leurs systèmes de sanctions divergent. Ainsi le maître peut manifester le plus vif plaisir dans telle ou telle activité d'enseignement. Par exemple, lorsqu'il fait des mathématiques, il se peut que les élèves soient irrités par une démonstration de plaisir. S'ils ont le sentiment que le maître recueille des joies considérables dans ce domaine où ils ont eux-mêmes l'expérience de désagréments considérables, il est peu probable qu'une forte imitation puisse se greffer là-dessus.

Naturellement, il conviendra de définir avec le plus grand soin ce que l'on entendra par imitation. L'entendra-t-on dans le sens d'une reproduction d'activités apparemment non pertinentes comme la façon de marcher du maître, ou sa façon de s'habiller, ou même sa façon de se moucher ? Par exemple il est fort possible que ce qui sera imité soit beaucoup plus le plaisir pris à une activité donnée, que cette activité elle-même. Par conséquent, l'imitation du plaisir pris à cette activité, fera peut-être que les élèves réaliseront une activité de type tout à fait différent de celle manifestée par le maître. A ce stade de la recherche, ceci ne représente évidemment qu'une hypothèse.

Il reste à examiner un aspect proprement méthodologique concernant l'effet sanctionnel des notes attribuées au cours de l'expérience. Il s'agit de savoir si ces notes constituaient, pour le maître, comme pour les élèves d'ailleurs, de vraies notes ; étaient-elles perçues comme telles par ces derniers ?

Nous rappelons que la procédure utilisée au cours de l'expérience est à peu près identique à celle pratiquée habituellement en classe ; c'est donc la réplique d'une situation scolaire réelle. Toutefois, ces notes n'étaient évidem-

ment pas prises en compte pour le «classement mensuel» des élèves. A cet égard on peut effectivement admettre qu'elles n'étaient pas de «vraies» notes au sens habituel.

Cependant, on n'avait rien dit aux élèves du destin de leurs «actes» et ils pouvaient supposer que, d'une façon ou d'une autre, leur conduite même de jugement pouvait être jugée et notée. Il conviendrait de poursuivre la recherche sur cette voie, d'amender alors le dispositif pour d'autres expériences dans lesquelles les élèves seraient des examinateurs effectifs, responsables de notes réelles qui seraient alors explicitement considérées comme véritables sanctions.

Cette expérience a par ailleurs permis de mettre en évidence un phénomène important ; les individus utilisent différemment un système d'appréciation, et donc de sanction , selon qu'ils sont ou non impliqués par ce système. C'est ainsi que nous avons pu relever une certaine propension à la «générosité» lorsque les sujets ont d'abord commencé par appliquer ce système à leurs propres devoirs.



REFERENCES BIBLIOGRAPHIQUES

Derghal. M. (1985) : Emprise inspiratrice et structure d'action : étude de psychologie sociale théorique et expérimentale. Paris. Laboratoire de psychologie sociale.

Derghal. M. (1992) : Approche psychosociale des processus d'apprentissage : esquisse d'une théorie des sanctions. *Psychologie*, n° 3, 27 - 38.

Noizet. G. (1961) : Etude docimologique sur la correction de l'écrit du baccalauréat. *B.I.N.O.P.* 17, 257 - 267.

Noizet. G. ; Bonniol. J. J. (1969) : pour une étude de docimologie expérimentale. *Bulletin de psychologie* ; 22, n° 276, 782 - 787.

Noizet. G. ; Caverni. J. P. (1978) : *Psychologie de l'évaluation scolaire* - Paris - P. U. F.

Pieron. H. (1963) : *Examens et docimologie* Paris - P. U. F.

Rosenthal. R. ; Jacobson. L. (1971) : *Pygmalion à l'école*. Paris ; Casterman.