

dominants), tracer le modèle de ce que l'on pouvait légitimement attendre de la réalité évaluée. On voit que si l'évaluateur peut décider que, dans tel cas, telle attente sociale est prioritaire et doit donc être un élément central du référent, il n'a pas, en tant qu'évaluateur, à juger de la pertinence des différentes attentes sociales. Son travail est postérieur au choix proprement axiologique : par exemple, au choix des finalités qui président à l'organisation d'un projet de formation. Il consiste à articuler avec cohérence les choix nécessaires dans les trois principaux "espaces" concernés par la mise en œuvre de son activité.

Car la multiplicité des questions que soulève la démarche d'évaluation s'ordonne si l'on prend en considération trois grands espaces de positionnement. L'évaluateur est le plus souvent préoccupé d'abord par les questions d'ordre technique : comment s'organiser ? Quels outils utiliser ? Quand prendre des informations ? etc. Or, nous avons pu montrer que les choix qui s'effectuent dans l'espace "méthodologique" sont déterminés par les choix faits, d'une part, dans l'espace de l'usage social, où des fonctions pourront être privilégiées en rapport avec le contexte décisionnel (niveau de gestion et type de décision à prendre) ; et, d'autre part, dans l'espace des intentions où s'organisent les stratégies en fonction de buts qui peuvent concerner l'objet évalué (à mesurer, à apprécier, ou à interpréter), ou la nature du rapport "évaluateur-évalué". Dans ces conditions, assurer l'évaluation consistera principalement :

- à s'interroger sur le ou les jeux auxquels on se livre ; à voir ce qui sépare, ou au contraire réunit, ceux auxquels, en tenant compte de ses propres valeurs et préférences, on aimerait jouer, et ceux auxquels on est contraint par le contexte institutionnel et social;
- à s'interroger sur les pièges propres à chacun, afin de tenter de ne pas en être victime.

— à développer, à partir de là, des stratégies cohérentes dans le cadre des choix effectués.

Il nous reste à montrer comment on peut s'efforcer d'atteindre ces deux derniers objectifs.

DEUXIÈME PARTIE

Assurer l'évaluation

Assurer l'évaluation : des questions de l'évaluation aux problèmes des évaluateurs et des évalués

Assurer c'est, au sens propre, mettre en état de sécurité et de confiance, à l'abri du danger. En quoi l'évaluation a-t-elle donc vraiment besoin d'être affirmée ? Contre quoi est-il urgent de la protéger ? En affirmant la nécessité d'assurer l'évaluation, ne cédon-t-nous pas, tout d'abord, aux facilités du discours par lequel on défend sa discipline, auxquelles succombent assez facilement professeurs et inspecteurs ? L'évaluation, "discipline" jeune et encore fragile, aurait besoin d'être confortée et défendue contre certaines tentatives d'empêtement. C'est sur ce mode de conflit de territoire que sont souvent vécus, par les enseignants, les problèmes d'horaires. Dès qu'un projet visant à mieux adapter les rythmes scolaires aux rythmes individuels conduit à envisager, dans le cadre d'une redistribution des enseignements, une légère diminution de l'horaire dévolu à telle ou telle discipline, la mobilisation est prompte. Personne ne voulant jamais abandonner une seule heure, les disciplines établies finissent toujours par triompher de tous les assauts. Elles continuent à bien se porter... tandis que sont malades les élèves qui subissent un emploi du temps aussi démesuré qu'inadapte. De même qu'un enfant "dominant", comme l'ont bien montré les travaux d'Hubert Montagner¹, marque son emprise en occupant certains lieux privilégiés, espaces clés dans l'espace socialement vécu de la crèche,

1. Cf., en particulier, *L'enfant et la communication*, Paris, Stock, 1978.

de même les disciplines s'efforcent de conquérir des positions fortes dans l'espace-temps scolaire, pour manifester leur existence et leur valeur. Dans cette logique, les élèves deviennent, tous ... des "dominés". Il ne s'agit en aucun cas, pour nous, de participer à un tel mouvement de promotion d'une discipline qui s'accomplirait au détriment de ceux au service de qui elle devrait être. La question, précisément, de savoir s'il fallait constituer l'évaluation en objet séparé, en "science" ou discipline autonome, a été posée récemment dans le cadre d'une réflexion sur la formation des maîtres à l'évaluation². Mais, en se critiquant sur le problème de "l'entrée" (entrée par la didactique ou entrée par l'évaluation ?), le débat sur la formation risque de s'enliser, une nouvelle fois, dans les problèmes stratégiques d'occupation de l'espace, en oubliant l'essentiel. Or l'essentiel nous paraît être, dans les questions de formation, l'intérêt des formés. Quels sont les besoins auxquels doit répondre la formation ? Quels sont les problèmes qu'elle peut contribuer à résoudre ? Ce qui revient à poser sous une nouvelle forme le problème du modèle de fonctionnement de l'activité d'évaluation. A quoi, et ici plus précisément à qui doit servir l'évaluation ? Car telle est notre hypothèse de travail : *l'évaluation est une activité qui doit s'exercer au profit de ceux sur qui elle s'exerce* (par exemple : les élèves), ou *qui sont concernés par l'objet sur lequel elle s'exerce* (dans le cas, par exemple, d'une méthode pédagogique). Cette hypothèse priviliege un modèle de fonctionnement : *l'évaluation contribue à la résolution de certains problèmes vécus par ses acteurs ou ses agents*. Vouloir la mettre au service de ceux-ci conduit donc d'abord à s'interroger sur ces problèmes.

Dans les activités d'enseignement ou de formation, comme dans tout processus enseignement/apprentissage, il y a deux grands types d'acteurs : les formateurs et les formés. Il s'agit d'aider ceux-ci à surmonter les difficultés qui leur sont propres en tant que formés. Tel est, en simplifiant, le seul problème des formateurs. Il est donc nécessaire, pour pouvoir "assurer l'évaluation", de savoir ce que requiert le formé, dans la situation d'être-formé, du point de vue de l'évaluation. Quelles sont donc les exigences de cette situation ? Quelles sont les conditions d'une formation réussie ?

Du point de vue social de l'activité d'évaluation, nous avons distingué au moins trois grands jeux possibles : celui de la gestion didactique (espace des apprentissages), celui de la gestion institutionnelle (espace du dialogue parents/enseignants/élèves), celui de la gestion sociale (espace de l'articulation école/société). Le formé peut ainsi être envisagé d'un triple point de vue.

- en tant qu'apprenant ;
- en tant qu'élève subissant une action de socialisation au sein d'une institution éducative où il "dialogue" avec les partenaires de l'acte éducatif ;

² Cf. "Savoir évaluer pour mieux enseigner. Quelle formation des maîtres ? Textes réunis par Monique Thivierge et Philippe Pennexotin. Cahier n° 26. Service de la recherche sociologique. Genève, 1988.

– en tant qu'individu destiné à s'insérer dans un réseau socio-économique, et construisant ce qui fera sa valeur sociale (ou sociale).

L'évaluation doit-elle se mettre au service de l'apprenant, de l'élève qui se socialise, ou du futur agent économique ? Il est clair que la première réponse est : si possible, au service des trois, puisqu'il ne s'agit jamais que d'une seule et même personne ! Mais nous avons vu que les logiques selon lesquelles s'organisaient les jeux différaient selon les contextes décisionnels. C'est ainsi que le jeu social exigera l'élimination de ceux qui ne résistent pas aux épreuves, de plus en plus difficiles et formelles (comme en témoigne le rôle décisif de sélection joué aujourd'hui par les mathématiques), imposées par un système scolaire qui trie les élèves en fonction d'exigences d'ordre social (enjeux de perpétuation d'une stratification sociale) ou technico-économiques (besoins de main-d'œuvre propres à une société industrielle avancée). L'élève en perdition en première année de classe préparatoire n'a pas à attendre de secours de l'évaluateur qui n'est là que pour trancher (évaluation-couperet) en séparant ceux qui suivent et ceux qui ne suivent pas, et qu'il convient d'éliminer. Le jeu institutionnel est moins cruel. Il n'exige que la saisie et la transmission ritualisées d'informations selon un rythme commandé par les temps forts du découpage du temps scolaire (les conseils de classe précédant chaque fin de trimestre, par exemple). La fonction de communication sociale qu'il assure n'est au fond qu'une fonction auxiliaire pour l'orientation/sélection autour de laquelle s'ordonne le jeu social, ou pour la fonction d'aide au développement individuel qui organise – ou devrait organiser – le jeu pédagogique. La logique de ce dernier est une logique d'aide à l'apprentissage. Ainsi, dans l'espace de son usage social, l'évaluation peut-elle concourir à trois grandes fonctions :

- orientation / sélection du futur producteur économique ;
- communication réalisée entre les "partenaires" sociaux d'un acte éducatif ;
- facilitation des apprentissages des formés.

La première règle pour qui veut évaluer rigoureusement est de savoir quel est, dans chaque cas ou chaque circonstance, le jeu dominant.

S'il s'agit de trier et de sélectionner, qu'on le fasse ! Mais à condition, d'une part, que ni les sélectionneurs ni les sélectionnés ne soient dupes en pensant faire autre chose que ce qu'ils font ; et, d'autre part, qu'on ait défini de façon claire les critères de tri. Dans certaines disciplines, le concours d'agrégation ne retiendra qu'un candidat sur dix. Les candidats éliminés étaient-ils les moins aptes à assurer un travail de enseignement ? S'il s'agit de communiquer de l'information utile aux partenaires de l'acte éducatif, la rigueur exige que l'on sache d'abord de quel type d'utilité il est question. Qu'est-il important de dire, et donc, en vue de quoi ? Assurer l'évaluation, c'est jouer le plus "proprement" le jeu dominant. Mais en fonction de quoi un jeu est-il dominant ? C'est poser le problème des droits et devoirs de

l'évaluateur, de ses obligations de "service". A quoi sert-il donc ? Sa fonction est déterminée socialement et institutionnellement. S'il est fonctionnaire, ses obligations sont définies par le législateur, qui traduit la volonté générale en textes de loi dont s'inspirent des notes et instructions. C'est dans ce cadre que l'enseignant/évaluateur doit orienter, communiquer, et faciliter des apprentissages. Mais chaque chose à sa place et en son temps. La première nécessité est de ne pas tout mélanger, et d'accomplir, en toute clarté, chacune de ces tâches.

Nous ne prendrons pas, toutefois, cette dernière distinction de fonctions pour fil conducteur de la suite de notre travail. Nous pourrions tenter, en effet, de dire successivement comment accomplir avec rigueur chacune d'entre elles. Mais nous considérons – deuxième hypothèse de travail – qu'il y a, en droit et en fait, une fonction dominante pour les enseignants ou les formateurs : celle de faciliter les apprennissages. Et nous l'affirmons en considérant ce qui nous paraît constituer l'essence de leur activité. Il nous faut donc définir plus précisément ce que signifie enseigner ou former.

Nous avons tenté de concevoir l'évaluation dans le contexte général de l'action humaine. Il était nécessaire de baliser l'immense domaine des jugements évaluatifs pour mieux comprendre et effectuer les actes d'évaluation qu'exige le processus éducatif. C'est dans ce cadre précis que nous resterons désormais. Mais, une nouvelle fois, de quoi parlons-nous ? Du travail des éducateurs, des enseignants, et/ou des formateurs ? Le processus éducatif se réduit-il, comme le laisseraient penser les premières lignes, au processus enseignement/apprentissage ? Puisque la première obligation de l'évaluateur est d'opérer en toute clarté, essayons de commencer en étant clair sur notre champ d'étude.

Nous proposons de distinguer des activités d'éducation, d'instruction, et de formation. Le terme d'éducation est celui qui recouvre le plus large champ d'activités. L'éducation est l'action exercée sur un individu ou un groupe d'êtres humains en vue d'infléchir ou de modeler leur comportement. Cela suppose que l'individu ou le groupe soient modifiables et capables de développement. On pourrait dire en ce sens qu'éduquer, c'est intervenir dans le développement d'autrui pour le rendre conforme à certaines normes. Ce qui fait voir que le premier problème pour l'éducateur sera celui de la légitimité des normes qu'il impose. Cette action qui, comme Guy Avanzini l'a bien montré³, est multidirectionnelle et ne vise pas un usage précis, concerne plutôt des devoir-être. L'enseignement (ensemble des activités dites d'instruction) est une action systématique ayant pour objet d'aider à s'approprier des outils intellectuels, dont ces outils privilégiés que sont les connaissances. L'enseignement est ainsi ordonné à promouvoir et à faciliter des apprentissages. Au sens large, l'apprentissage est la construction (à la fois

l'opération et son résultat) de nouveaux modèles de comportements. Ces modèles peuvent être vus comme des outils, utiles à l'individu dans son dialogue avec l'environnement physique et humain. L'enseignement qui s'effectue dans les établissements scolaires a pour objet de faciliter les apprennissages particuliers que l'école a pour fonction première de gérer. Par exemple : lire, écrire, calculer, à l'école élémentaire. Ces apprennissages concernent les trois principaux axes du développement : intellectuel (construction des savoirs), mais aussi moteur et sensoriel. Enfin, on parlera de formation lorsque s'exerce un ensemble cohérent d'actions en vue de conférer à des sujets une compétence précise et prédeterminée⁴.

	DÉFINITIONS	CONTENUS ET DOMAINES CONCERNÉS
ÉDUCER	Intervenir dans le cours d'un développement, pour l'infléchir dans un sens jugé souhaitable.	Savoir-être Relations interhumaines
ENSEIGNER	Aider un élève à s'approprier les outils intellectuels propres à une discipline. Gérer les apprentissages de l'élève.	Savoirs Connaissances
FORMER	Conférer une compétence à la fois précise et limitée, et pré-déterminée.	Savoir-faire en situation Activités sociales

Son organisation même autour des "disciplines" qui structurent son espace-temps montre que l'Ecole a pour mission première d'instruire ; que là est son "essence". Ce qui permet de comprendre, et fonde, notre deuxième hypothèse de travail : la tâche centrale de l'enseignant est d'assurer les apprennissages scolaires. Cette tâche est "éducative", dans la mesure où il y a intervention dans un développement. La pluralité des disciplines marque le souci d'accroître la polyvalence de la personne. Mais, d'une part, le choix des disciplines est le fruit d'une première décision sociale concernant les directions à privilégier, et traduit (et fige en même temps) un consensus sur un programme commun et minimal de développement. Et, d'autre part, on vise des capacités et compétences qui, sans correspondre à des savoir-faire

3. Guy AVANZINI, *Introduction aux sciences de l'éducation*, Toulouse, Privat, 1987, p. 136.

4. Guy AVANZINI, *Introduction aux sciences de l'éducation*, op. cit., p. 136.

en situation (compétences professionnelles) sont relativement précises et limitées, puisqu'on devrait théoriquement pouvoir en dresser la liste. Cette tâche, d'un autre côté, est donc préformatrice, dans la mesure où les outils acquis par apprentissage seront des pré-requis nécessaires pour une authentique formation visant un type déterminé d'insertion socio-économique. Ces considérations nous autorisent, pensons-nous, à affirmer que, la facilitation des apprentissages étant la tâche essentielle des enseignants, l'évaluation scolaire a pour fonction première de contribuer à une bonne gestion de ces apprentissages. Et nous posons, en conséquence, qu'assurer l'évaluation dans le domaine de "l'éducation" (c'est-à-dire à l'Ecole, autour du pôle central de l'enseignement), c'est lui permettre d'assurer efficacement cette fonction. Mais, d'une part, la pluralité des stratégies possibles d'aide à l'apprentissage, organisées autour de tâches jugées prioritaires (apprentissage, contrôle, facilitation), et la diversité des fonctions repérées à l'intérieur du seul espace de jeu "pédagogique", montrent qu'il y aura plusieurs façons de contribuer à une bonne gestion des apprentissages. Et comme, d'autre part, la prise en compte du rapport qui unit l'évaluateur à son objet nous a appris que les pratiques d'évaluation prenaient finalement un sens différent selon la "philosophie" dominante de l'évaluateur, qui se comportera comme un expert, un juge ou un philosophe, il peut être judicieux de tenir de repérer les pièges auxquels on s'expose, plus particulièrement, dans chacun de ces trois grands cas. Chacune de ces philosophies prédisposant, par ailleurs, à certains usages sociaux du discours évaluatif, cela nous permettra d'aborder aussi les problèmes propres aux fonctions de dialogue socio-institutionnel et d'orientation-sélection, en les situant dans un cadre qui devrait en faciliter l'élucidation. Ainsi, pourront être mises en évidence quelques conditions d'émergence d'une évaluation mise au service des partenaires du processus "d'instruction", et permettant aux uns d'accomplir avec rigueur leur travail de facilitateur d'apprentissage, et aux autres de bénéficier d'informations utiles pour leur propre développement. Nous devrions alors être en mesure de poser en toute clarté le problème des outils et des dispositifs.

Premier moment

Savoir repérer et déjouer les pièges auxquels on s'expose selon sa "philosophie"

Les pièges de la parole objective : la problématique de l'estimation et les impasses de l'objectivisme ; vraies et fausses notes

Assurer l'évaluation, c'est donc, tout d'abord, permettre au praticien de "l'enseignement" de faire correctement ce qu'il a à faire en ce domaine. Or l'institution attend de lui qu'il fournit des informations fiables sur le travail des élèves, sur leurs résultats. Ces informations se présentent encore aujourd'hui, de façon dominante, sous la forme de notes. La place tenue par la note dans la vie scolaire est assez prodigieuse. La vie institutionnelle des enseignants est, comme celle de leurs élèves, scandée par des notes chiffrées : note administrative et note pédagogique décident de l'avancement d'adultes qui sont, sous cet aspect, placés dans la même situation (infantilisante ?) que les enfants ou les adolescents dont ils sont... les "maîtres" ! On pourrait parler d'un véritable impérialisme, et d'un impérialisme réducteur. Car un seul chiffre est censé tout dire, alors qu'originellement la note est notation, au double sens de représentation par un symbole, et de brève communication écrite. Que reste-t-il du message communiqué lorsque le symbole est coupé de ce qu'il a pour rôle de représenter ? Peut-on condenser un discours, même bref, en un seul signe ? Il faudra décoder, c'est-à-dire interpréter, un signe sibyllin qui lui-même exprime – ou devrait exprimer – une première interprétation ! L'usage d'un tel moyen de communication ouvre la porte à tous les malentendus. Une première nécessité, pour diminuer les risques d'équivocité, serait donc de créer les conditions d'un langage commun en dotant le locuteur et l'allocuté d'un référentiel permettant d'éviter de

trop fortes distorsions dans la compréhension du message, c'est-à-dire dans la mise en correspondance du signifiant (la note chiffrée) et du signifié (ce qu'il y a à communiquer). La note n'est utile que si les partenaires du dialogue dont elle est l'instrument possèdent le même lexique définissant les domaines de référence où prendront signification les observations qu'elle condense¹, et précisant les significations attachées, dans chaque cas, aux différentes valeurs ; ce qui fait voir que le problème du bon usage de la note déborde largement la question docimologique.

Pendant longtemps, le paradigme docimologique a dominé les travaux portant sur l'évaluation. Dans la mesure même où l'on réduit l'évaluation à la notation, on voit dans la recherche de procédures "scientifiques" de notation le principal problème de l'évaluateur. Le terme de docimologie a été proposé par H. Piéron en 1922, à partir des termes grecs : *dokimē*, qui signifie épreuve, examen ; et *logos*, parole, raison, d'où : discours scientifique. La docimologie correspond à la tentative de construire une parole "scientifique" sur les épreuves et les examens. Elle peut être définie comme science des examens ou, d'un façon plus large, comme l'étude scientifique des procédures d'examen, en particulier des systèmes de notation². Le paradigme docimologique, qui a été dominant des années 1940 aux années 1960, a connu un premier élargissement important lorsqu'on s'est intéressé, dans les années 1970, au comportement des examineurs.

Mais cet élargissement ne s'est pas traduit par un changement significatif dans la perspective. Il s'agit toujours de déterminer les conditions d'une évaluation plus objective. On veut pouvoir saisir plus rigoureusement l'objet évalué. Mais quel est cet objet ? Le simple fait de poser une telle question va nous permettre de comprendre que la réalisation de l'intention même qui sous-tend le projet de produire des notes-mesures objectives implique paradoxalement le passage d'une problématique de l'objectivité à une problématique de la pertinence. On ne peut en effet attendre une amélioration significative des pratiques d'un simple perfectionnement des procédures et systèmes d'estimation. Ni la connaissance des mécanismes en jeu, ni celle des causes de distorsions ne sont suffisantes pour assurer l'évaluation estimative.

La première condition pour dire "objectivement" le "poids" de la réalité, selon l'ambition propre à la philosophie de la parole objective en laquelle prennent sens, à l'évidence, les pratiques de notation, est de triompher du piège de l'objectivisme, en s'interrogeant d'abord, et toujours, sur ce qu'il faut connaître objectivement. C'est à ce prix que deviendra plus rigoureux le jeu qui se joue dans l'espace du dialogue social parents/élèves,

et que ce jeu pourra être mis au service du jeu "pédagogique" de la facilitation des apprentissages en lequel nous a paru s'exprimer l'essence du travail des enseignants. Telle est la principale leçon qui se dégage d'une étude des difficultés de la notation, qui présentera, par ailleurs, l'intérêt de nous permettre de nous approprier quelques acquis fondamentaux de la recherche en docimologie.

Vraies et fausses notes : la fin d'un mythe (panique dans la notation)

Pour le sens commun, noter, c'est mesurer. Or, nous le savons, mesurer signifie assigner un nombre à un objet en saisissant celui-ci à travers un système d'unités qui s'y appliquent, et qui est valable pour tous les objets du même type. Et évaluer revient à dire ce qu'il convient de penser d'une réalité en fonction d'une grille de lecture particulière, et construire à cet effet. Le principal risque de l'évaluation estimative – par défaut de mesure – est d'oublier cette distinction, et de prétendre alors produire un discours qui soit la mesure des choses. Si l'on prétend pouvoir, en quelque sorte immédiatement, mesurer un objet, alors que l'on n'est pas véritablement en mesure de pouvoir le faire, pour des raisons qui tiennent à la nature de l'objet (qui ne se réduit pas à ce qui est donné de lui dans l'espace), ou à son "ambiguïté" ("l'objet" réellement visé – par exemple la compétence du producteur – étant distinct de l'objet saisi – la copie produite) ; alors donc que l'on ne peut que tenter de dire (produire un discours, nécessairement équivoque) ce que l'on voudrait mesurer (saisir dans un système de symboles univoques), on est presque condamné à ne plus savoir ce que l'on dit. Tel est le piège de l'*objectivisme*, dont aurait dû faire prendre conscience la critique docimologique.

Car cette critique a mis en évidence un fait massif, qui est l'absence de fiabilité des procédures de notation. Rappelons trois exemples devenus classiques.

– Une comparaison des moyennes des notes attribuées, dans une même matière, par différents jurys de baccalauréat, entre lesquels les candidats avaient été répartis au hasard (relatif) de l'ordre alphabétique, montre que ces moyennes fluctuent d'une façon importante : par exemple, de 5,81 à 9,06 pour l'épreuve écrite de mathématiques, discipline traditionnellement considérée comme donnant lieu aux estimations les plus objectives (session de juillet 1955, pour treize jurys)... Les jurys ne corrigeant pas de la même façon, les divergences étant plus importantes encore pour les épreuves orales. Conséquence importante : les pourcentages de candidats admis par

1. Cf. B. MACCARIO, "L'évaluation", art. cité, p. 365.
2. Cf. PIÉRON, *Examens et docimologie*, Paris, PUF, 1963.

les différents jurys fluctuent en proportion (par exemple, de 31 à 53 % dans les cas ci-dessus³).

– Les expériences de multicorrection ont mis en évidence que des examinateurs différents ne corrigeaient pas de la même façon. Une expérience réalisée en 1967 dans les académies de Lyon, Clermont et Limoges, met en lumière l'importance des fluctuations. Corrigées par 150 professeurs différents enseignant en classe de troisième, trois compositions françaises obtiennent des notes qui varient de 4,5 à 13,75 pour la première, de 2,5 à 12,5 pour la seconde et de 5,5 à 17,5 pour la troisième. Pour trois devoirs de mathématiques, les notes fluctuent respectivement de 0,5 à 11,5, de 11,5 à 20, et de 3,5 à 11,5⁴.

– Enfin, les notes attribuées varient également dans le temps, ce qui conduit G. de Lansheere à parler "d'infidélité" du correcteur,⁵. Une recherche que cite Pierton, portant sur des compositions d'histoire, le met en relief. Quinze copies ont reçu la même note moyenne. On les soumet à quinze autres correcteurs : on obtient quarante notes distinctes. Douze mois plus tard, les mêmes correcteurs mettent des notes différentes. Et sept mois plus tard, les notes diffèrent encore⁶.

Du constat à la recherche des causes

De tels résultats n'étonneront pas si l'on songe à la multiplicité des causes possibles de biais. On pourrait distinguer trois sources principales de distortion, correspondant à des facteurs individuels, sociaux, ou tenant au processus même d'évaluation.

– L'absence de fiabilité des notes tient, en premier lieu, au rôle joué par la "subjectivité" de ceux qui les attribuent. On peut voir ici tout ce qui sépare l'évaluation par défaut de mesure de l'évaluation-interprétation. Pour J. Ardoino et G. Berger, alors que le contrôle repose sur "l'idéal de l'interchangeabilité parfaite des contrôleurs", l'évaluation est le propre d'un sujet que l'on ne peut remplacer sans modifier, par là même, le discours évaluatif⁷. L'évaluateur, de ce point de vue, est irremplaçable. Or, déployer la non-fiabilité des notes, c'est vouloir que les correcteurs "n'évaluent" pas, et regretter qu'ils ne soient pas interchangables ! Pour que le "contrôle" des connaissances soit assuré et juste, il conviendrait que des correcteurs diffé-

rents accordent les mêmes notes aux mêmes productions. S'ils ne le font pas (ce qui vient d'être constaté), c'est, pour une part, parce qu'ils ne contrôlent pas la dimension subjective de leur réaction au produit qu'ils notent.

Déjà, les notes peuvent varier selon l'humeur, la disponibilité, ou l'état de fatigue de celui qui les attribue. Mais elles sont aussi fonction de ce à quoi chacun sera plus sensible, au-delà de l'impression immédiate produite. Ainsi tel correcteur accordera une grande importance à la maîtrise de la langue, même pour un devoir de mathématiques ; tel autre à la présentation, etc. La diversité des critères rend possibles de multiples combinaisons⁸. Les notes, enfin, ont une dimension idéologique. Elles dépendent de certaines représentations propres à l'évaluateur : de l'image, par exemple, qu'il se fait de son rôle et de la façon dont il doit l'assurer⁹. Les jurys des grands concours de recrutement d'enseignants se distinguent en ce sens par leur volonté de maintenir le niveau, et leur refus d'une évaluation "laxiste" qui ferait fi des impératifs d'exigence et d'excellence...

Mais la réalité offre parfois bien des surprises, et les comportements ne sont pas toujours conformes aux représentations. Une expérience citée par G. de Lansheere montre que les correcteurs n'ont pas nécessairement une conscience objective de leur degré d'indulgence ou de sévérité. La moyenne des notes attribuées à quatre devoirs fut de 12,4 pour les correcteurs se croyant sévères, et de 11,6 pour ceux qui se pensaient indulgents¹⁰.

– En second lieu, les variations déplorées peuvent être d'origine "sociale", dans la mesure où, comme l'ont très bien montré P. Bourdieu et J.C. Passeron, le jugement professoral peut se guider, inconsciemment, sur des signes sociaux. Cela est parfaitement visible à l'oral de certains grands concours, où l'appréciation des savoirs et savoir-faire des candidats est "contaminée" par la prise en compte d'un "système de manières caractéristiques d'une position sociale". L'examen est aussi une épreuve de reconnaissance sociale. La société savante et cultivée va reconnaître (et s'agrégérer...) les siens en se fondant certes sur des indicateurs de compétence, mais aussi sur des indices exprimant, dans le maniement de la langue, le maintien, les gestes, le vêtement, et la mimique, un rapport à la culture et au savoir propre à un groupe social¹¹.

Au-delà de ces mécanismes sociaux de reconnaissance ou d'exclusion qui jouent lors du "face à face" notateur/noté, on peut décrire également comme facteur "social" de biais, en un sens plus large, l'enracinement ou inscription "géographique" de l'acte de notation. Les notes sont relatives au contexte, et expriment des traditions ou particularités locales, propres à un

3. Cf. Maurice Beuchelin, "Problèmes d'évaluation", in *Traité des sciences pédagogiques*, t. IV : *Psychologie de l'éducation*, Paris, PUF, 1974, p. 213-214.

4. Cf. *Didactologie et examens*, Lyon, CRDP, mars 1980, 44 p.

5. Cf. G. de LANDSHEERE, *Évaluation continue et examens*..., op. cit., p. 35.

6. Cf. H. PIERTON, *Évaluations et didactologie*, op. cit., p. 15.

7. J. ARDOINO, G. BERGER, "L'évaluation, comme interprétation", art. cité, p. 122 et 123.

8. Cf. B. MACCIONI, *Ce que valent nos enfants*, op. cit., p. 43.

9. Cf. Marie-Claire DAUVIS, "Quelques propos pour engager un débat" in De KETELÉ Éd., *L'évaluation : approche descriptive en perspective*, Bruxelles, De Boeck, 1986, p. 215.

10. G. de LANSHIERE, *Évaluation continue et examens*..., op. cit., p. 30.

11. P. BOURDIEU, J.-C. PASSERON, *La reproduction*, Paris, Minuit, 1970, p. 149-150.

établissement, selon son histoire, ou à une région. Et les évaluateurs adaptent, par nécessité, les notes qu'ils attribuent aux caractéristiques socio-culturelles de leur public. Ils tiennent compte du niveau de leurs classes.

— Enfin, l'analyse psychologique de l'acte d'évaluation, avec en particulier les travaux de G. Noizet et J.P. Caverni, a mis en évidence l'existence de causes qui sont en rapport avec cet acte lui-même, dans sa dimension de tâche psychologique.

Pensant que toute modification de procédure risque de rester vaine tant que l'on ne saura pas à quoi sont dues les divergences entre évaluateurs, G. Noizet et J.P. Caverni se sont attachés à connaître la façon dont l'expert s'y prend pour former ses jugements, et à décrire les mécanismes psychologiques que les évaluateurs-experts mobilisent dans l'exécution de leur tâche. Ils sont ainsi parvenus à isoler quelques-unes des caractéristiques du fonctionnement cognitif de l'évaluateur, en mettant en évidence les effets produits sur la notation par certaines variations provoquées dans la situation d'évaluation. Entre autres, cinq facteurs de variation ont été identifiés, et étudiés expérimentalement :

- l'ordre des productions dans le lot évalué ;
- les informations dont on dispose sur les producteurs ;
- la "place" des indices prélevés dans une production appréhendée nécessairement de façon diachronique ;
- la combinaison ou l'articulation d'indices relevant de plusieurs dimensions (ex : style, orthographe, calcul, raisonnement) ;
- la nature du langage utilisé pour énoncer l'évaluation.

Les résultats les plus intéressants concernent les deux premiers facteurs, qui se sont révélés être d'importantes sources de variation pour la note.

D'un côté, phénomène important dans la mesure où les productions évaluées font en général partie d'un lot (par exemple : une série de copies), la recherche conduit à mis en évidence des effets d'ordre, en relation avec ce que J.-J. Bonniol a décrit comme phénomène d'ancre¹². Certaines copies, fortement marquées (très bonnes ou très mauvaises) jouent le rôle d'ancre, modifiant la perception des travaux suivants. D'où, des *effets de contraste* : une même production est jugée plus faible quand elle suit une ancre haute ou forte, et plus forte quand elle suit une ancre basse. Les notes dépendent, pour partie, de la place de la production jugée dans le lot.

D'un autre côté, les notes peuvent fluctuer selon les informations que possède l'évaluateur sur le producteur. On parle alors d'*effet de source* : une même production est jugée différemment selon la qualité supposée ou connue du producteur ; ou d'*effets d'assimilation* : assimilation d'une éva-

luation à celles qui ont été attribuées antérieurement au même élève. Dans tous les cas, l'élève réputé fort est mieux noté ; et, inversément, les progrès de l'élève réputé faible auront du mal à être perçus, ce qui amène Noizet et Caverni à conclure que les procédures d'évaluation ont leur part de responsabilité dans l'échec scolaire.¹³ L'effet de source est à rapprocher de l'*effet de halo*, quand il y a contamination d'aspects différents : influence exercée, par exemple, par des indices d'ordre "affectif" (l'allure du candidat) ou formel (la présentation de la copie), sur l'apprehension des aspects techniques (exprimant la compétence visée). Les analyses de la *reproduction* décrivent, en ce sens, un halo social. Et l'effet d'assimilation, du mécanisme général de stéréotypie, qui peut être défini ici comme une contamination des résultats (de Landsheere). Ainsi la note peut-elle être "ajustée", par souci de minimiser un risque d'erreur et d'éviter la dissonance entre deux séries d'informations :

- aux performances antérieures de l'élève (effet d'assimilation) ;
- à son statut scolaire (effet de source) ;
- à son statut social (phénomène de catégorisation).¹⁴

Contre le mythe de la note vraie

Tout cela interdit-il aux évaluateurs de tenter de progresser vers une notation plus juste et plus fiable ? Le constat de l'importance des causes de biais doit-il conduire à désespérer ? L'expertise est-elle hors de portée ?

A première vue, il n'y a pas d'obstacle fondamental qui rendrait caduc le projet de progresser vers une évaluation conforme au modèle de l'expert, capable de dire la réalité telle qu'elle est, selon l'intention caractéristique de la philosophie de l'évaluation par défaut de mesure. Au contraire, la connaissance des facteurs de distorsion ne rend-elle pas justement possible la recherche de procédures susceptibles de ne plus les faire jouer ? Des remèdes pourraient être mis en œuvre, en regard de chaque grande cause de biais, comme le montre le tableau 7 (constats, causes, remèdes), p. 98.

De fait, plusieurs procédures sont devenues d'utilisation sinon courante, du moins fréquent. Les plus connues et utilisées tendent à lutter contre les divergences entre évaluateurs. On procède par modération, a priori, ou remédiation, a posteriori. A posteriori, il est possible soit de procéder à une multicorrection, censée diminuer l'incertitude de la note ; soit à un traitement statistique harmonisant les échelles en leur donnant même moyenne et dispersion. A priori, on peut prévenir les distorsions en harmo-

13. G. NOIZET, J.-P. CAVERNI "Les procédures d'évaluation ont-elles leur part de responsabilité dans l'échec scolaire?", *Revue française de pédagogie*, 62, janvier-mars 1983, p. 7-14.

14. Cf. G. de LANDSHEERE, *Évaluation continue et examens... op. cit.* p. 41.

nissant les points de vue et les attentes par la notation en commun de quelques devoirs, ou, mieux, déterminer un barème analytique précisant les différentes lignes de lecture de l'objet évalué (ou les différents "sous-objets" constitutifs de la production), et la valeur ou l'importance à attribuer à chaque aspect pris en compte. Cette dernière procédure a le mérite de commencer à déplacer l'accent du problème de la "vérité" d'une note à celui de sa signification.

Car il est remarquable que chacun des remèdes envisagés ait suscité des critiques. Les normalisations a posteriori ne constituent qu'un remède formel, et ne font en rien disparaître la dimension aléatoire de l'appréciation de ce qui serait le niveau intrinsèque de la production, dimension que ne gommant pas davantage les épreuves communes et normalisées... sauf à supprimer l'évaluateur lui-même. A cet égard, l'utilisation des QCM (questionnaires à choix multiples), ou des tests de connaissance, est exemplaire. La notation est devenue automatique : il suffit de compter les bonnes réponses. Mais que note-t-on, en dehors de la capacité de restituer des connaissances, avec en plus une part d'aléa avec le QCM, puisque l'on pourra toujours répondre au hasard ? Résout-on le problème en supprimant l'évaluateur ? Le triomphe de la recherche de l'expertise serait-il dans la disparition de l'expert ? Sans doute, le QCM présente-t-il un triple mérite qui lui assure une "réputation d'objectivité"¹⁵. Il procède par échantillon de questions et constitue ainsi un sondage significatif. Il supprime pratiquement les ambiguïtés de formulation des sujets et problèmes. Il permet de parvenir à un résultat unique, le calcul du score étant, en principe, toujours automatisable. Toutefois, il présente également des inconvénients importants, comme l'a souligné Louis Legrand¹⁶. Il est inadéquat pour évaluer des activités "supérieures", telles que les activités de synthèse, ou des comportements sociaux. Il ne permet pas davantage d'évaluer des créations complexes. Enfin il renforce la tendance "à l'abstraction formelle", et contribue ainsi à faire disparaître, comme objets d'évaluation – et par là même, comme objet d'enseignement – des capacités dont l'acquisition devrait être un objectif essentiel pour la formation, telles que les capacités d'observer, d'émettre des hypothèses, de vérifier ces hypothèses. Sans doute, comme le pensent Noizet et Caverne, la pertinence des objections soulevées par les procédures d'évaluation automatisées peut-être apparaître très inégale. Mais le fait qui nous paraît mériter réflexion est l'apparente disparition de l'évaluateur. En réalisé, la procédure du QCM sépare en deux parties distinctes la tâche de celui-ci. L'évaluateur-notateur, celui qui attribue péniblement une note, disparaît bien. Il peut être remplacé par une machine. Mais dans la mesure seulement où, en amont, l'évaluateur-concep-

15. G. NOIZET, J.-P. CAVERNE, "Les procédures d'évaluation...", an. cité, p. 8.
16. LOUIS LEGRAND, "Les problèmes de l'évaluation scolaire", in *Les Amis de Sévres, L'évaluation formative*, n° 3, septembre 1983, p. 9, 10.

TABLEAU 7. – Les difficultés de la notation

CONSTATS	Causes	Remèdes possibles
Divergences inter-médiatrices (entre évaluateurs)	• Particularités des points de vue - Préférences personnelles (ex : QCM) • Suppression du notateur	Cause(s)
Divergences inter-individuelles	• Particularités des points de vue - Choix idéologiques • Multicorrélation	Remèdes possibles
Difficultés locales	• Préférence du contexte - Traditions et particularités locales • Attention accordée aux manières révélant Epreuves communes et/ou normalisées	Attention accordée aux manières révélant une situation de classe
Survalorisation de la forme	• Préférences de « Classe » - Au « niveau » de la classe - aux caractéristiques socio-culturelles	Adaptation
Relativité de la note	• Effets de source ou d'assimilation (produits par le producteur) - à ce que l'on sait du producteur, d'autres - à la situation du produit parmi d'autres - Evaluation « isolée » Evaluation « anonyme »	Evaluation « anonyme » Evaluation « isolée » Produits

teur-de-la-tâche-d'évaluation a soigneusement délimité les champs de questionnement, et précisé les critères qui constitueront l'ossature du barème analytique. Ce n'est que parce qu'il y a eu ainsi tout un travail portant sur l'objet de l'évaluation, sur ce qu'il faut évaluer, sur les tâches qui permettent de manifester que l'on possède les qualités recherchées, que la notation peut devenir machinale. Car le score ne traduit jamais que le degré "d'atteinte" de l'objectif. La seule conclusion est ici, comme l'écrivent Noizet et Caverni, "qu'il convient d'ajuster la procédure d'évaluation à l'objectif poursuivi".¹⁷ On pourrait dire que, de ce point de vue, l'objectivité est dans la centration sur l'objectif, et que la vérité de la notation ne peut être trouvée que dans la construction de procédures critériées – ce qui marquerait, d'une certaine façon, une limite radicale pour la philosophie de l'estimation. Mais, plutôt, l'expert apparaît moins comme celui qui dit les choses comme elles sont (perspective objectiviste) que comme celui qui sait de quoi il parle.

Ainsi, l'intérêt majeur du QCM nous paraît être d'induire une réflexion sur les contenus et les tâches d'évaluation et, à travers celle-ci, sur les objectifs et les méthodes. Et de faire comprendre finalement que le problème essentiel pour l'évaluateur-notateur n'est pas de trouver la vraie valeur d'un produit, mais de pouvoir communiquer, d'une façon claire, un jugement porté sur le degré de réalisation d'un projet précis. La quête de l'objectivité par la recherche de la vraie note est illusoire et mystifante, comme le confirme une réflexion sur les expériences de multicorrection.

Car, quel est le sens de la multicorrection comme procédure de remédiation ? Si l'on voit dans la notation une entreprise de mesure, il est logique de penser que, comme par exemple dans le domaine des sciences physiques, on pourra lutter contre l'imprécision et diminuer l'incertitude en multipliant le nombre des mesures concrètement effectuées. Le coefficient de fidélité d'une série de mesures augmente avec leur nombre. En utilisant une formule qui permet de prévoir cette augmentation, on a pu calculer le nombre minimal de "mesures" – ici, de correcteurs – nécessaires pour obtenir ce qui serait la "vraie" note, des mesures supplémentaires ne faisant plus varier significativement la précision. On obtient le tableau suivant¹⁸.

Composition française	78
Version latine	19
Anglais	28
Mathématiques	13
Dissertation philosophique.....	127
Physique	16

Retenant le problème, J.-J. Bonniol¹⁹ montre tout d'abord qu'en tenant compte de certaines données ayant permis d'obtenir ces nombres et en particulier du nombre initial de correcteurs dans l'enquête effectuée en 1933-1936, il conviendrait de multiplier les résultats par 6. Ce qui donnerait, par exemple, 78 correcteurs en mathématiques et 762 en philosophie ! Mais surtout, en critiquant les hypothèses de travail sur lesquelles repose une telle recherche, il montre que celle-ci n'a guère de sens. En effet, adopter la théorie de l'erreur variable pour les sciences physiques implique que l'on accepte deux hypothèses :

- a) l'erreur (l'incertitude, l'imprécision) tient à l'opération de mesure. Elle provient des conditions de mise en œuvre de l'instrument de mesure (constitué, ici, par le correcteur).
- b) de même qu'un objet n'a qu'une longueur, ou qu'une épaisseur, une copie n'a qu'une valeur, sa "valeur vraie", que la mesure (la note) a pour fonction de donner, de faire connaître.

Or ces deux hypothèses sont particulièrement contestables, et pour la même raison, qui tient à la nature de l'objet évalué, qui lui interdit d'être "estimé" au sens strict. La première pose que le correcteur effectue une mesure, dont il est l'instrument. La mesure est, par essence, vérifiable. Pour être mesuré, un objet doit être définissable sur une seule dimension. Si l'objet a plusieurs dimensions, chacune doit être isolable, et sera estimée séparément (longueur, épaisseur, etc.). Et l'on ne pourra estimer l'ensemble, c'est-à-dire l'objet considéré en même temps sous toutes ses dimensions, que si celles-ci peuvent être ramenées à une nouvelle échelle, unique. Si bien, conclut J.-J. Bonniol, que l'on ne pourra estimer d'une façon vérifiable un objet complexe que si chacune des dimensions à prendre en compte est : 1 : isolable ; 2 : pourvue d'une échelle numérique ; 3 : combinable avec les autres selon une seule règle de composition ; 4 : la nouvelle dimension unique obtenue pouvant être munie d'une échelle numérique. Or la nécessité même de construire des barèmes analytiques montre qu'une production scolaire est multidimensionnelle. Elle n'est pas un objet "univoque" ; et des dimensions telles que le style, la précision, l'élegance, ne peuvent être munies d'une échelle numérique. Que peut-on faire avec un tel objet ? L'évaluer, c'est-à-dire l'apprécier par référence à une échelle, non de mesure, mais de valeur.

Est-ce à dire que l'évaluation-estimation n'est qu'un leurre ? Oui, si l'on croit pouvoir mesurer, si l'on pense qu'une note est une valeur numérique dont l'attribution est vérifiable. Non, si l'on se réfère à l'intention qui la caractérise, d'être objectif, et qui prend tout son sens quand on la rapproche du souci de "parler vrai" qui domine la première philosophie décrite du rapport évaluateur/évalué. Car il faut comprendre qu'être objectif, ce

17. G. NOIZET, J.-P. CAVEANI, "Les procédures d'évaluation...", art. cité, p. 9.
18. Cf. H. PIERON, *Éléments et doctrinologie*, op. cit., p. 23.

19. J.-J. BONNIOL, "La fidélité en question, objet d'étude doctrinologique", *Commentarius*, 19, 1976, p. 63-78.

n'est pas, ici, saisir "scientifiquement" un objet mesurable, mais porter un jugement assuré sur la valeur de cet objet, envisagée *d'un point de vue objectivable*, c'est-à-dire explicitable et communicable. Une production scolaire n'est pas un coffre contenant une quantité déterminée d'or. Sa valeur n'est pas une qualité intrinsèque. L'hypothèse de travail (b) est essentiellement. C'est pourquoi il faut se délivrer du mythe de la vraie note, et abandonner un paradigme qui conduit à des conclusions proprement délinantes, comme en témoignent les nombres trouvés pour les correcteurs nécessaires.

Du bon usage de la note pour ne pas être victime du paradigme docimologique

D'une problématique de l'objectivité à une problématique de la pertinence

L'évaluateur ne doit donc pas se laisser mystifier par une analogie trompeuse avec le physicien qui se sert d'une balance pour peser son objet. Et le principal obstacle à l'évaluation objective n'est pas la subjectivité du correcteur. Ou pas exactement. Car s'il importe de ne pas être emporté par son humeur ou ses préférences idéologiques, il importe surtout de savoir ce que l'on fait, de clarifier le jeu auquel on se livre. Comment l'évaluateur-instrument d'évaluation peut-il devenir un instrument fiable ? La fiabilité est le caractère de ce qui est digne de confiance. D'où provient donc la confiance que l'on peut accorder à l'individu-instrument ? Sans doute, de sa probité, et de son sens de la justice. La notation ne doit pas devenir le moyen d'un règlement de comptes ; et, nous y reviendrons, il ne faut pas méanger travail de police et travail d'évaluation. Mais la confiance repose aussi, et surtout, sur la qualité des informations transmises. Qu'est-ce donc qu'une information de qualité ? Celle qui a un sens pour le locuteur comme pour l'allocuté. Ce qui, implique, nous l'avons vu, la possession d'un lexique commun, et l'explication des domaines de référence où prennent sens les observations faites. Le premier travail de l'évaluateur est de s'assurer de l'existence d'un tel lexique, ou de contribuer à sa construction en explicitant, précisément, les domaines de référence en jeu. Car évaluer n'est pas peser un objet que l'on pourrait isoler sur le plateau d'une balance. C'est apprécier cet objet *par rapport à autre chose* que lui.

Toute évaluation est une occasion de prendre et de donner de l'information. La question importante est de savoir quel genre d'information, et

pour faire quoi. Le tableau des fonctions (prédictive, formative, sommative) pourra ici servir de guide utile. Car on pourra dire que l'on n'utilisera pas la même balance selon la fonction assignée à l'acte concret d'évaluation, qui peut s'ordonner selon des logiques différentes, en fonction d'objectifs différents. Autrement dit, pour évaluer "objectivement", il convient :

- de déterminer la question précise à laquelle l'évaluation a pour objet d'apporter des réponses précises. Par exemple, concernant une dissertation, la question : "Est-elle bonne ?" n'est pas opératoire. "S'organise-t-elle pour apporter une réponse argumentée à la question cachée dans le sujet ?" est déjà une question plus précise, permettant de fournir de l'information utile à l'auteur du travail. On pourrait imaginer, à partir de là, un outil de communication de l'évaluation présentant, sur autant de lignes qu'il y a, en fait, de dimensions à estimer, ou de domaines de référence, les questions que l'évaluateur s'est posées à propos du produit, et les réponses apportées. Car il convient deuxièmement :
 - d'expliquer la réponse apportée à chaque question, en la justifiant. (Exemple : Question : "Avez-vous su discerner et exprimer le plus clairement possible, avec le moins de mots possible, la question contenue dans le sujet ?" Réponse : "Non, il n'est pas possible de surigner dans votre introduction une phrase qui exprimerait cette question".)

Il est clair qu'un tel travail, qui conduit à une utilisation dynamique de ce qui pourrait constituer, par ailleurs, un barème analytique (la grille d'apprehension ou de lecture du produit), équivaut simplement, selon la démarche décrite dans notre première partie, à expliciter des critères (lignes de lecture correspondant à des domaines de référence), et à rechercher, pour chacun d'eux, des indicateurs.

L'évaluation est pertinente quand, en fonction du contexte décisionnel, elle précise les axes de questionnement du produit et s'organise pour apporter des éléments de réponse fondés à des questions claires. L'évalué, sachant de quoi il est question, pourra alors vraiment bénéficier de ce qui lui est dit. Voir le critère sous forme de question permet de comprendre comment *l'évaluation est dialogue*.

Pour un dialogue utile et une évaluation résolument informatrice

L'intérêt d'une évaluation pertinente est d'être véritablement informative. L'évaluation est le moment et le moyen d'une communication sociale. Il faut en tirer toutes les conséquences. Nous avons commencé, en posant le problème de la pertinence, à déterminer les conditions d'une bonne communication. L'évaluation doit toujours fournir à l'élève des informations qu'il puisse saisir et qui lui soient utiles. Si la note apporte une information compréhensible et utile, pourquoi s'en priver ? Mais souvent l'information

apportée est indirecte, et se situe dans une zone de non-dit, le dialogue se déroulant sur le registre de l'implicite. C'est ce qu'a très bien fait saisir Yves Chevallard 20 en décrivant la "négociation didactique". L'acte d'évaluation est un moment fort dans la pièce aux aspects divers qui constitue le procès d'enseignement, à la dimension quelquefois dramatique au sens restreint du terme. Dans la vie, souvent conflictuelle, de la classe qui, "dans une perspective fonctionnelle d'ensemble", peut être saisie comme une "totalité de fonctionnement et de signification", la note attribuée a valeur de message. Elle traduit, le plus souvent, un "marchandage" entre l'enseignant, qui s'efforce de faire travailler et avancer les élèves, et ceux-ci, qui s'emploient à retarder, autant qu'ils le peuvent, la marche de la classe. Dans ce combat où les uns et les autres, condamnés à vivre ensemble, doivent sans cesse "négocier", la notation devient une arme aux mains des enseignants, ou plus précisément une carte que l'on abat dans les moments de difficile "discussion". Il y a un "langage de la note", qui n'a de sens précis que dans "l'intimité de la classe". Ainsi, mettre 9,5, c'est refuser de mettre 10. Le professeur ne dit pas vraiment : "Cette copie vaut 9,5" (langage de la valeur), mais plutôt : "Je l'achète 9,5" (langage du négoce).

Entre l'idéal mystifiant de la note-mesure, qui ne dit plus rien parce qu'on ne sait ni quel est l'objet "mesuré", ni quel est le système d'unités qui s'y applique, et la réalité quotidienne de la note-message-élément de négociation, qui ne dit jamais franchement ce qu'elle a à dire et, ne parlant que par ce qu'elle ne dit pas, parle mal et en dit toujours trop ou trop peu, ne peut-on, en prenant acte du fait que l'évaluation est toujours communication, essayer au moins de parler clair en faisant du discours évaluatif un message qui ait du sens pour celui qui le reçoit ?

Pour un déplacement de l'intérêt du produit au producteur

Le mythe de la vraie note naît, entre autres, d'une attention excessive accordée au produit. N'avons-nous pas centré notre réflexion, dans ce chapitre, sur l'évaluation des productions scolaires ? Nous ne nous délivrerons du mythe qu'à condition de passer d'une problématique centrée sur le produit (qui serait à "peser"), à une problématique centrée sur le processus de production, à connaître et améliorer, et finalement sur le producteur, à aider. Pour être un acte de communication utile, l'évaluation doit renouer le fil avec le producteur, et lui dire quelque chose de sa production qui lui permette de progresser vers de meilleures productions. A travers le produit, c'est l'activité de production, et le producteur lui-même, en tant qu'il

construit des capacités et compétences, que vise l'évaluation, comme lecture orientée, animée et dynamisée par un questionnement. Celui-ci, dans le cadre d'une stratégie d'aide, concernera tout autant l'élève-producteur (caractérisé par un ensemble de capacités ou compétences), que son activité de production (mise en œuvre des compétences) ou les produits eux-mêmes (résultat de cette mise en œuvre). Qu'y a-t-il à dire d'une dissertation ? Si l'on y retrouve ou non les critères d'un produit "acceptable" (ex. discrimination et compréhension du sujet, organisation d'un parcours argumenté vers la réponse, etc.), qui manifestent la maîtrise de compétences que l'enseignement s'efforce de développer chez l'élève (ex. compréhension, analyse, expression, synthèse).

Ni mesure, ni négociation : mais un dialogue permanent comme élément central d'une cogestion des apprentissages

Sans doute, est-il désormais clair que l'évaluation, comme estimation, ne réalisera pas son ambition de dire les choses comme elles sont en devenant capable de saisir l'essence d'un produit et d'en dire la valeur intrinsèque qu'elle pourrait mesurer, mais en s'efforçant de parler vrai, selon la première des "philosophies" décrisées du rapport social. Et parler vrai, ce n'est pas dire la vérité d'un objet, mais donner à un sujet des informations fiables – dignes de confiance –, sur quelques aspects de ses productions importants et significatifs par rapport aux apprentissages qu'on l'aide à réaliser, aux compétences qu'on l'aide à construire. L'évaluation, délivrée de la tentation objectiviste, peut alors nourrir un dialogue permanent qui permettra à l'élève-apprenant de cogérer, véritablement, ses apprentissages, avec l'enseignant-facilitateur. En s'appuyant sur des informations qui vont l'éclairer, le guider, l'encourager : en l'aïdant à analyser son activité ; en attirant son attention sur ses forces et ses faiblesses ; en lui permettant de voir où il en est. Toute information qui va dans ce sens est utile.

Faut-il, dans ces conditions, conserver ou supprimer la note ? Le seul problème est de savoir quelle information elle véhicule. Si elle permet de se situer – mais alors c'est le rang qui importe, et non le score²¹ ; si elle sera à matérialiser un niveau de maîtrise²², pourquoi s'en priver ? Mais si elle témoigne simplement de la persistance du mythe de la vraie valeur, ou n'est que l'instrument de visées qui n'ont plus rien à voir avec l'évaluation – maintien de l'ordre ou imposition de sa force – il faudrait mieux ne plus en faire usage.

20. Yves CHEVALLARD, "Vers une analyse didactique des faits d'évaluation", in De Riedel éd., *L'évaluation, approche descriptive ou prescriptive ?*, Bruxelles, De Boeck, 1986, p. 31-59.
21. Cf. Ateliers lyonnais de pédagogie, *Du han usage de la maturation*, 2^e édit., Méthode Barat-Becker, numéro spécial, novembre 1987 (16, quai Claude-Bernard, 69007 Lyon).

22. On trouvera un exemple de cet usage avec un outil proposé dans le chapitre 8.

valeur du comportement évaluée par rapport au comportement cible s'expose peut-être à des pièges encore plus redoutables, non seulement parce qu'il risque toujours de se tromper dans son appréciation, en attribuant, par exemple, à une absence de compétence, une erreur due à une tout autre cause. Mais aussi parce que l'évaluation appréciative avec modèle prédéterminé est exposée à une double dérive, comme on pourra le voir en remarquant que la "philosophie" du juge peut accompagner de façon élective deux des quatre conceptions du rapport social décrites plus haut, pour lesquelles elle constitue un arrière-plan en quelque sorte naturel. En effet, et de façon assez paradoxale, le discours du pouvoir, qui permet à l'évaluateur d'en imposer à l'évalué, et le discours du devoir, qui exprime la nécessité de venir en aide à celui-ci, s'incarnent dans des démarches orientées par la philosophie du juge, et se traduisent par des pratiques où l'on confronte l'individu évalué à un modèle idéal qui préexiste à cette confrontation. Le recours à un référent prédéterminé pour dire la valeur d'une production ou d'un produit prédispose ainsi à deux dérives : la dérive autoritariste, particulièrement à craindre dans l'espace du dialogue social parents/enseignants/élèves, et la dérive techniciste, dont sont menacées, comme pratiques, la pédagogie par objectif (PPO), ou même l'évaluation formative. Comment ces dérives se caractérisent-elles ? Et comment l'évaluateur pourra-t-il les éviter ?

Les limites de la parole appréciative :

la problématique de l'appréciation et les dérives autoritariste et techniciste

S'il ne veut pas connaître les égarements auxquels peuvent conduire la prétention de savoir et de dire les vraies nature et valeur d'un objet ou d'une personne (tentation essentialiste), et la conviction que l'objectivité et la subjectivité varient en sens inverse, et qu'il n'y a, en conséquence, de salut que dans la mesure (tentation "quantophénétique")¹, l'expert doit devenir modeste et comprendre que, pour être valide, une évaluation n'a pas avant tout à être "exacte", mais à se construire en cohérence avec les intentions qui sont à son origine. Que dire alors du juge ? Quels sont les risques auxquels s'expose celui qui, ayant saisi que les ambitions de l'enseignement étaient d'abord d'ordre qualitatif, puisqu'il s'agit pour l'élève de construire des compétences, de développer des capacités, ou de s'approprier des outils, avant d'effectuer des "performances" éventuellement mesurables, se propose précisément d'apprécier le comportement de ses élèves par rapport à un modèle prédaté-miné qui sert de critère-cible ? On voit bien que la tentation de la mesure n'est plus guère à redouter, puisqu'à la limite un objectif est atteint ou non, une compétence est présente ou non. Traduit en chiffres, cela donnera 0 ou 1, comme dans la logique de l'ordinateur. Cependant, le juge qui dit la

Évaluation et pouvoir d'évaluer

La dérive autoritariste

Jean-Marie Barbier, nous l'avons vu, observe que si le droit d'évaluer est présenté comme une caractéristique principale de l'évaluation des actions de formation par l'ensemble des acteurs impliqués dans leur réalisation, en réalité l'exercice de ce droit est fonction des rapports de pouvoir existant dans les actions elles-mêmes². Le pouvoir d'évaluer ne se distribue pas de manière égale et de façon équitable. Selon son degré de participation aux processus de définition des objectifs et de détermination des indicateurs, le formé peut n'être qu'un simple objet qui subit l'évolution, ou au contraire un sujet actif qui construit les conditions d'émergence d'un jugement pertinent par rapport à ses besoins et à ses préoccupations, et dont il pourra finalement bénéficier pour son propre développement.

Il est certain que le cas le plus défavorable est celui où lui échappe non seulement la construction du référent, mais même la connaissance de

1. Selon le terme (quantophénétique) utilisé par le sociologue américain Sorokin pour dénoncer les abus de la quantification des phénomènes sociaux et l'usage intempestif de la statistique.

2. J.-M. BARBIER, *L'évaluation en formation*, op. cit., p. 238 et 249.

celui-ci. Le cas où l'évaluateur juge par référence à des "lois" qui ne sont ni écrites ni communiquées, et en utilisant des critères qui restent implicites ou, en tout cas, demeurent cachés. C'est alors que peut naître, et de façon justifiée, un sentiment d'injustice fondé sur la perception du caractère éminemment arbitraire d'une telle pratique. Comment un élève qui ne voit pas ce que l'on reproche exactement à son travail, ou qui se rend compte que les critères de jugement de l'enseignant fluctuent d'un "contrôle" à l'autre, pourrait-il ne pas être tenté de conclure que le maître évalue "à la tête du client"? Ce n'est pas, alors, la subjectivité du maître qui est vraiment en cause, et fait obstacle, mais l'incertitude quant aux critères d'appréciation. Car quand on connaît ces critères, même s'ils peuvent paraître contestables, et si leur bien-fondé peut être discuté, au moins dispose-t-on de points de repère permettant de s'orienter et d'organiser son travail. Ainsi l'évaluation avec référent prédéterminé est-elle une démarche particulièrement contestable lorsque, paradoxalement, fait défaut ce référent, c'est-à-dire lorsque celui dont le travail est jugé n'a aucune idée précise de ce qui détermine et fonde le jugement. Et la première nécessité, pour qui veut jouer correctement ce jeu, est en conséquence de faire connaître ses critères. P. Bourdieu et J.-C. Passeron ont décrit l'action pédagogique comme violence symbolique, en tant qu'elle est imposition, par un pouvoir arbitraire, d'un arbitraire culturel³. Il y a violence symbolique quand un pouvoir impose des significations comme légitimes en dissimulant les rapports de forces qui, de fait, ont présidé à cette imposition. On pourrait parler en ce sens de violence douce, d'autant plus perfide qu'elle dissimule sa réalité, lorsque l'évalué est contraint, en recevant la note ou l'appréciation qu'on lui propose, d'admettre par là même la légitimité d'une opération de jugement qui repose sur des critères cachés ou obscurs, alors que cette absence de référent clair crée les conditions d'une contestation de cette légitimité. Y a-t-il pouvoir plus arbitraire que celui qui dissimule son caractère arbitraire en occultant soigneusement ce qui le rend arbitraire? Telle est la première, et sans doute plus insidieuse, forme de ce que nous proposons de nommer dérive autoritariste.

On pourrait ici objecter que la non-communication des critères d'évaluation manifeste davantage une négligence, sans doute coupable, qu'elle ne traduit une volonté de pouvoir ; et qu'il conviendrait plutôt de déplorer une insuffisance à laquelle il devrait être facile de remédier, que de condamner un autoritarisme qui n'est pas évident. Il faut certes se garder de généraliser hâtivement, et de condamner trop vite. Mais le risque n'est-il pas réel de se complaire dans une situation floue, et de la pérenniser, parce qu'elle procure l'avantage d'exercer un pouvoir sans avoir à le justifier, de juger sans avoir à donner ses raisons ? N'y a-t-il pas lieu d'être inquiet quand, au début de nombreux stages de formation continue d'enseignants, l'affirmation qu'il faut "permettre à l'élève de s'approprier les critères de l'enseignant" est

assez massivement rejetée ? Ne refuse-t-on pas ce qui serait vécu comme une violence exercée par l'élève sur le maître — quelque chose lui étant "arraché" — pour pouvoir continuer à exercer en toute impunité la violence subtile du jugement souverain ? Sans doute, pour la plupart des enseignants, une telle dérive reste-t-elle au stade de la tentation. Mais il convient de prendre conscience de cette tentation pour ne pas se laisser aller au plaisir "coupable", dénoncé par Patrice Ranjard, que l'on éprouve dans l'ivresse d'être "le maître absolu de ses notes" ; au plaisir du "Pouvoir avec un grand P", que procurera l'exercice du "jugement sans appel".

Pour Ranjard, ce plaisir est proprement infernal. La jouissance est telle qu'on ne peut la reconnaître. Ce serait reconnaître que, contrairement à ce que l'on prétend et proclame, on est moins l'humble serviteur du développement de l'élève qu'homme de pouvoir, à la recherche d'un "plaisir de mauvaise qualité mais sûr, garanti, quotidien". Celui qui procure "la toute-puissance de noter : un plaisir qui vient des enfers et qu'on ne peut regarder en face..."⁴. Pour éviter cette dérive, la première règle de l'évaluation avec modèle prédéterminé doit être de rendre transparent ce modèle.

Mais, lorsque le "modèle" est explicité et communiqué, tout risque n'est pas pour autant écarter. Car, alors, se pose la question de sa légitimité. La violence qui constitue l'imposition de normes de comportement ne se justifie que dans la mesure où ces normes peuvent être jugées, d'une façon ou d'une autre, légitimes. Autrement dit, on ne peut pas échapper au problème de l'évaluation de ce au moyen de quoi on évalue. L'évaluateur, qui n'est pas, en tant que tel, un prescripteur, doit cependant, pour ne pas exercer un pouvoir de prescription qui présenterait alors le double caractère d'être subreptice et arbitraire, pouvoir faire reconnaître le bien-fondé du référent qui lui sert de grille de lecture. On pourrait dire, en se référant à la théorie du double arbitraire, que la violence de l'arbitraire du mode d'évaluation (la forme particulière de ce coup de force qu'est un acte d'évaluation qui, même lorsque les choses se passent en douceur, n'en est pas moins une "violence symbolique" exercée sur le sujet évalué) sera d'autant plus durablement ressentie qu'apparaîtra arbitraire le contenu imposé (non plus la forme : conditions et circonstances de l'acte, mais la nature du modèle imposé). Plus contestable sera la norme, plus l'évaluation devra être contrainte dans ses procédures. Inversement, plus le référent sera perçu comme légitime, plus l'opération pourra être douce. Ainsi, une des conditions de possibilité de l'auto-évaluation est-elle l'appropriation par le sujet d'un référent absolument légitime à ses yeux.

Or le problème de la légitimité se pose différemment selon que l'acte d'évaluation se situe dans un contexte éducatif, instructif ou formatif. S'il s'agit de former, la norme sera fondée dans la mesure où elle exprime bien

3. P. Bourdieu, J.-C. Passeron, *La reproduction*, op. cit., p. 19.

4. Patrice RANJARD, *Les enseignants pénétrés*, Robert LAURE, 1984, p. 94 et 99.

une tâche fondamentale dans le champ professionnel pour lequel on forme. Il est légitime d'évaluer un enseignant à l'aune de ses capacités de communication si l'acte pédagogique est acte de communication. S'il s'agit d'instruire, le problème est double. La première question est celle de la légitimité de la discipline comme moyen de développement. Celle-ci étant admise, il faut affronter alors la question de la pertinence du modèle choisi pour juger des progrès de l'apprenant, comme nous le verrons en examinant le risque de dérive techniciste. S'il s'agit d'éduquer, la légitimation des normes n'est pas évidente, et le risque de dérive autoritariste devient fort.

Une première façon de surmonter ce risque serait donc de s'attacher plutôt à former ou à instruire qu'à éduquer : car, comme en ont témoigné certaines critiques de l'éducation scolaire particulièrement vives dans les années 1970, la volonté d'éduquer fait naître le soupçon de vouloir enfermer dans un moule, de vouloir dresser les individus à penser et à agir conformément à un modèle social dominant. Cette dénonciation de l'action normalisante de l'école s'est développée autour de quatre arguments⁵.

- le système scolaire est par nature autoritaire et répressif ;
- il opère un modelage arbitraire ;
- il sclérose l'individu ;
- il enferme dans le quantitatif.

Le quatrième argument qui exprime, à sa façon, le refus de l'évaluation-mesure, a été en particulier avancé par Carl Rogers et Ivan Illich. Pour C. Rogers, l'utilisation de ce qu'il appelle "l'horrible machinerie" des tests, examens et évaluations, repose sur l'erreur fondamentale qui consiste à croire en la possibilité d'une mesure extérieure du développement personnel. Or il serait "absurde" de vouloir mesurer l'apprentissage par des notes d'examens⁶. La croyance stigmatisée est un aspect de ce qu'Ivan Illich désigne comme "le mythe des valeurs étalonées", qu'il juge particulièrement dangereux. Car "les hommes qui s'en remettent à une unité de mesure définie par d'autres pour juger de leur développement personnel, ne savent bientôt plus que passer sous la toise"⁷. L'évaluation avec référent prédéterminé, en imposant ses normes, qui sont extérieures à l'individu, court ainsi le risque de créer des habitudes de soumission, chacun finissant par venir se glisser de lui-même à la place qui lui est assignée par la société. On voit comment ce dernier argument reprend et rassemble les trois autres.

Un système autoritaire éduque en fait à l'obéissance, l'individu subissant et finissant par accepter et admettre comme légitime un mode de jugement arbitraire qui le sclérose. Cette analyse critique repose sur deux conceptions concordantes du développement et de l'éducation. Un développement personnel est quelque chose d'irréductible, qui ne peut être rapporté à rien

d'autre. "C'est croissance, épanouissement unique d'une dissidence personnelle et, pourtant, ordonnée, qui ne se mesure à aucun étalon de comparaison."⁸ Et l'éducation a pour fonction d'accompagner le développement, et non de le figer ; de sauvegarder et faire fructifier les potentialités individuelles, et non de les étouffer sous un masque social.

Nous ne saurions avoir la prétention de dire ici en quoi peut consister une éducation respectant la nature, ni de montrer comment la socialisation peut être quand même libération individuelle⁹. Nous ne pouvons que reconnaître l'existence du risque de normalisation et inviter l'éducateur-évaluateur à s'interroger toujours sur l'intérêt et la valeur, pour un authentique développement du sujet évalué, de la norme à travers laquelle ce développement est apprécié. Par exemple, que vaut le modèle, qui paraît dominant aujourd'hui, de l'élève qui "gagne" parce qu'il travaille vite, dans l'abstrait, et surmonte avec bonheur les épreuves qu'on lui impose à un rythme accéléré ? Rapidité d'exécution, maîtrise formelle des langages abstraits, et surtout capacité de résister à la pression des examens, sont-elles les caractéristiques majeures d'un individu bien développé ? Il faut peut-être réfléchir avant d'entrer dans ce jeu, en tout cas avant d'en rajouter comme savent si bien le faire les défenseurs du "niveau". Lorsque manquent les points de comparaison, et alors que le public scolaire s'est transformé, quel sens peut avoir l'affirmation quasi obsessionnelle qu'il faut maintenir le niveau ? N'est-ce pas une façon d'exprimer son désir de maintenir les choses en l'ordre, et de fixer chacun à sa place ?

Le recours à un référent prédéterminé peut, certes, ainsi, permettre à l'évaluateur d'affirmer son pouvoir en l'exerçant de manière à contribuer à la conservation et/ou la reproduction d'un certain ordre des choses, en particulier d'une certaine stratification sociale. Tout comme l'usage de la "mesure" peut avoir pour effet de soumettre et de rendre docile. Toutefois, pas plus que la volonté de saisir "objectivement" n'est en soi condamnable, la question essentielle étant de savoir quelles sont, exactement, les propriétés "mesurées", l'usage d'un modèle prédéterminé ne transforme automatiquement l'évaluateur en agent de police sociale. Il faut se souvenir ici de la distinction opérée par J. Ardoino entre les formes, archaïque et moderne, du contrôle, pour s'interroger sur les conditions qui feront de l'évaluation un acte, non de police sociale, mais d'optimisation d'un fonctionnement, ce qui est l'ambition même de l'évaluation formative.

Mais, avant de voir comment celle-ci pourrait surmonter le risque, qui lui est propre, de dériver techniciste, nous pouvons, en prenant en considération cette ambivalence du contrôle, à la fois mieux saisir le mécanisme du

5. Cf. Charles HADJI, *Pédagogie et libération*, thèse de doctorat d'Etat, vol. I, p. 131-144.

6. Carl ROGERS, *Liberté pour apprendre*, Paris, Dunod, 1972, p. 260 et 230.

7. Ivan ILICH, *Une société sans école*, Paris, Le Seuil, 1971, p. 73.

8. Ibid.
9. On pourra consulter à ce sujet : Charles HADJI, *Pédagogie et libération*, op. cit., vol. 2.

piège qui guette ici l'évaluateur, et préciser les moyens de le surmonter. Car le contrôle est le cas limite de l'évaluation appréciative avec référent pré-terminé. Lorsque le référent peut être totalement "expliqué", lorsqu'il peut être rendu transparent, l'évaluateur ne se borne-t-il pas à vérifier que l'on trouve bien, dans la réalité concrète, ce qui est inscrit sur la check-list qu'est devenu le référent ? C'est pourquoi, tandis que l'évaluation, comme interprétation, est multidimensionnelle, produit de la compréhension et du changement, a un caractère démocratique, "autorisant" et dynamogène, le contrôle peut apparaître comme normalisant, hiérarchisant, dépossédant, sanctionnant et clôturant, et finalement conservateur¹⁰. Mais il est aussi, en tant que processus de vérification périodique de l'état d'un système, un temps indispensable dans la conduite de l'action. Il est bien nécessaire de contrôler l'existence des acquis ou la possession par l'apprenant des outils intellectuels que l'enseignant a pour tâche centrale de faire acquérir. Ardoino et Berger remarquent qu'avec les systèmes d'auto-contrôle, "le contrôle disparaît... comme acte extérieur spécifique pour s'intégrer totalement dans le processus de production lui-même."¹¹ Mais cela ne constitue-t-il pas, d'un certain point de vue, l'idéal pour le processus d'auto-production par l'apprenant de ses propres capacités ? Tout dépend de l'usage fait du contrôle : simple moyen de vérification externe d'une présence attendue; ou acte donnant prise sur un système en évolution et permettant, par la connaissance qu'il fournit des phases de cette évolution, de la guider, de la conduire. Si le risque de dérive autoritariste est réel, toute maîtrise n'est pas condamnable. La difficulté pour l'évaluateur est de savoir devenir un maître... auxiliaire, qui mette son activité de contrôle au service du développement de l'apprenant. Comment ? Dans le cadre, justement, d'une pédagogie du développement :

- a) en refusant les référents passe-partout, pour construire et proposer des référents dynamogènes. Autrement dit, en construisant, pour chaque contrôle, le référent adéquat :
 - en tant qu'il constitue une cible vraiment "désirable" pour l'apprenant, car correspondant pour lui à une étape supérieure du développement;
 - en tant qu'il n'est pas hors de sa portée relativement à ses capacités et compétences actuelles;

- b) en rendant ce référent transparent pour l'apprenant (appropriation par celui-ci des critères de réussite et, nous le préciserons, de réalisation);
- c) en permettant à l'apprenant de s'approprier enfin les résultats du contrôle, c'est-à-dire en lui communiquant des informations utiles, parce qu'utilisables.

Cela conduit, à l'évidence, vers une pédagogie par objectifs et une évaluation formative, ou "formatrice". On pourrait dire, pour en terminer avec ce point, que l'utilité de l'évaluation appréciative avec référent pré-terminé dépend d'abord de la valeur de celui-ci comme outil de développement. La difficulté réside davantage dans le choix d'une cible appropriée que dans la "mesure" de l'écart à la cible.

Objectifs et "formation" ; la dérive techniciste

Le repérage des pièges de la parole à prétention objective, et une première analyse de la problématique de l'appréciation conduisent à la même conclusion : pour surmonter le risque de dérive autoritariste, comme pour éviter de s'enfermer dans l'impasse de l'objectivisme, il convient de déterminer de façon rigoureuse l'objet de l'évaluation (savoir ce que l'on veut "mesurer" dans un cas, définir la compétence cible dans l'autre).

L'évaluateur est donc tenu de préciser ses objectifs de formation (ou d'éducation, ou d'instruction), puisqu'il s'agit d'en apprécier les résultats. Le jeu de l'évaluation appréciative implique la formulation d'objectifs pédagogiques, comme le jeu de l'évaluation estimative implique la recherche d'une cohérence objet/critère/indicateur. Lorsqu'il s'agit d'évaluation scolaire, étant entendu que l'on joue en priorité le jeu pédagogique de l'aide à l'apprentissage dans ce que nous avons appelé l'espace du processus didactique, l'objectif pédagogique a pour raison d'être d'exprimer, de faire apparaître, l'objet que l'on veut "mesurer". Ici se rejoignent les deux philosophies de l'expert et du juge. La philosophie de l'expert ne peut se réaliser, dans le cadre du jeu "social" rappelé ci-dessus, que par le moyen de la formulation précise d'objectifs, vers quoi conduit naturellement la construction de barèmes analytiques lorsque sont définis, non plus simplement les critères du bon produit, mais ceux du bon "producteur"... autrement dit les compétences que l'action de formation doit se donner comme cible. Cela permet de comprendre la faveur dont jouit, du moins dans les discours, l'évaluation formative.

Car les pratiques ne correspondent pas toujours aux discours, et l'on parle sans doute beaucoup plus de l'évaluation formative qu'on ne la met en pratique. Si nous parvenons à comprendre pourquoi, nous verrons mieux en quoi consiste le second obstacle majeur auquel se heurte la parole appréciative. Le problème est, en effet, de savoir comment faire de l'évaluation critériée un véritable outil au service du développement. L'examen de cette question va venir confirmer, tout d'abord, ce que nous venons d'entrevoir : si

¹⁰ Yves HARVEY, "Le contrôle, cet obscur objet du désir", *POUR*, 107, p. 117-118.
¹¹ J. ARDOINO, G. BERGER, "L'évaluation, comme interprétation", art. cité, p. 123.

n'est plus possible, lorsque l'intention est de gérer des apprentissages, de traiter séparément problèmes d'évaluation et problèmes de gestion didactique.

La didactique peut être définie comme "l'étude des procédures d'enseignement et de formation, et la recherche des plus pertinentes". Son objet est "d'éclairer les voies qui conviennent à la fois à celui qui enseigne et à ceux qui sont enseignés"¹². L'accent mis sur la nécessité d'une évaluation formative marque le glissement significatif d'une problématique autonome de l'évaluation à une problématique de l'évaluation comme auxiliaire de l'apprentissage. L'évaluation formative est centrée sur la gestion des apprennissages. Elle cherche à guider l'apprenant pour facilier ses progrès. Elle se veut procédure pertinente de formation. C'est pourquoi la réflexion sur ses problèmes et ses limites nous aidera à préciser ce qu'est l'activité didactique. La notion d'évaluation formative reposant sur trois concepts clés : ceux de critère, de diagnostic, et de régulation, nous examinerons successivement les difficultés inhérentes à la réalisation des intentions qui s'expriment en chacun de ces concepts.

Difficultés de mise en œuvre d'une évaluation critériée

L'évaluation formative a pour premier objectif de permettre à l'apprenant de savoir ce que l'on attend de lui et de se situer en conséquence. Tel est l'intérêt majeur de ce que l'on a nommé pédagogie par objectifs. Un objectif est un "énoncé d'intention pédagogique" qui traduit en termes de capacité ou de compétence de l'apprenant le résultat escompté d'une séquence de formation¹³. Il exprime le changement observable attendu dans le comportement de l'élève. On dit qu'il est opérationnel ou opéronnalisé lorsqu'il précise (règle des trois "C") :

- quel Comportement sera observable ;
- dans quelles Conditions il sera observé ;
- les Critères qui permettront d'apprécier la réussite (niveau de performance acceptable).

Exemple 1 : Sans aide, sans documents, mais sans limitation de temps, l'élève sera capable d'écrire quatre lignes sans faute.

Exemple 2 : En utilisant son manuel de statistique et une calculette, l'étudiant en sciences de l'éducation sera capable de calculer sans erreur le coefficient de corrélation entre deux variables.

Il est clair qu'une telle formulation de ce que l'apprenti doit être en mesure de faire à l'issue d'une séquence de formation présente de nombreux avantages. "L'entrée" en pédagogie par les objectifs permet, entre autres :
– de prendre du recul par rapport à la pratique, et d'instaurer une réflexion concrète sur l'acte d'enseignement et ses conditions. Comme l'a exprimé J.-P. Astolfi¹⁴, un tableau d'objectifs est un instrument "d'autoscopie" qui renvoie aux enseignants une image distanciée de leur propre travail ;
– de planifier le processus éducatif par référence aux résultats escomptés, ce qui à la fois clarifie les perspectives et constraint à articuler les finalités (grandes intentions) et les contenus (savoirs savants et savoirs à enseigner) avec la pratique concrète ;
– de rendre les choses plus transparentes pour l'apprenant, qui voit ce que l'on attend de lui, ce qui rend possible l'établissement de contrats didactiques clairs.

Toutefois, il est clair aussi que la mise en œuvre d'une pédagogie par objectifs ne va pas de soi, et expose à plusieurs risques :
– risque de faire un usage directif et normalisant des tableaux d'objectifs,

– risque de formalisme : on peut traduire tout un programme, exprimer toutes ses intentions en termes d'objectifs répondant à la règle des trois "C". Est-ce un gage de réussite pédagogique ? Les listes les plus détaillées et les formulations les plus adéquates peuvent apporter à leur rédacteur la satisfaction d'avoir pensé à tout et d'avoir trouvé de belles formules... qui resteront lettre morte tant que l'on n'aura pas trouvé les moyens concrets de faire progresser l'élève. Or la définition précise de l'objectif ne dit pas comment il sera possible d'aider l'élève à l'atteindre. Le fait que l'évaluation soit devenue l'auxiliaire de la didactique ne lui confère pas, comme par magie, un pouvoir pédagogique ;
– enfin, risque de parcelisation en objectifs de plus en plus réduits. Le choix n'est-il pas, finalement, entre des objectifs trop généraux et des objectifs trop précis ? Comment trouver la juste distance entre une "programmation atomiste, objectif par objectif", qui ferait perdre de vue l'unité du développement souhaité et évalué¹⁵, et des objectifs qui, dans leur généralité, resteraient abstraits et ne seraient plus appréhendables par l'apprenant ?

Cette dernière difficulté nous met sur la voie de comprendre comment, paradoxalement, en formulant des objectifs en termes de compétence observable, on risque de rester à la surface des choses, et sans s'en rendre compte, ce qui constitue le premier aspect du piège "techniciste". La dérive techniciste, en effet, guette celui qui croit que l'acquisition d'une technique

12. Guy AVANZINI, "A propos de didactique : il n'y a pas de consensus", *Bulletin Univer-Sinon*, 606, 1, 1986, p. 3 et 9.
13. Daniel HAMELINE, *Les objectifs pédagogiques en formation initiale et en formation continue*, Paris, ESF, 1979.

14. In *Caillers pédagogiques*, 148-149.
15. J. CARDINET, *Enseignement scolaire et pratique*, Bruxelles, De Boeck, 1986, p. 75.

de formulation d'objectifs suffit pour résoudre à la fois les problèmes d'évaluation et d'apprentissage. Car la compétence de l'évaluateur ne se réduit en aucun cas au "savoir formuler un référentiel". Encore faut-il, et pouvoir fonder ce référentiel sur un modèle pertinent du développement, en particulier cognitif, et pouvoir fournir à celui qui s'efforcera d'atteindre la cible des informations susceptibles de l'aider à le faire.

Nous avons dit que les objectifs formulés dans le cadre de l'évaluation appréciative explicitait l'objet que s'efforçait de mesurer l'évaluation estimative. Mais il n'y a pas de correspondance ou d'adéquation automatique objet/objectif. Un objectif pédagogique peut correspondre à plusieurs objets possibles (fig. 6). Car les objectifs ne sont jamais que des indicateurs ou indices qui renvoient à autre chose, dont la maîtrise est effectivement visée à travers eux. On pourrait, à ce propos, distinguer trois types d'objets, en fonction de deux axes :

- un axe concret/abstrait (ou observable/inobservable) ;
- un axe fonctionnel/structural, séparant les démarches et les produits de la pensée.

Les objectifs peuvent renvoyer :

- 1) soit à des contenus : faits, lois, concepts, etc., qui sont décrits et répertoriés dans des programmes (ex : la symétrie orthogonale, la Révolution française) ;
- 2) soit à des opérations intellectuelles ou cognitives : activités mentales qui permettent au sujet d'apporter une réponse adéquate à une situation donnée (ex : reproduire, conceptualiser, explorer) ;
- 3) soit à des tâches socialement significatives, ou compétences sociales (ex : soigner, enseigner, etc.).

Le comportement cible décrit par l'objectif, par définition toujours observable, n'est ainsi, selon une belle expression de Jean Cardinet, que "le sommet apparent d'un iceberg difficilement contrôlable".¹⁶ La description du comportement attendu "n'épuise pas" tout ce que l'enseignant-formateur entend par maîtrise de l'objectif.¹⁷ Même lorsque l'objectif est centré sur une compétence sociale observable (savoir réparer un robinet qui fuit), celle-ci met en jeu des savoirs et savoir-faire qui ne sont jamais directement observables (savoir analyser une situation, explorer, mobiliser, etc.) et qui constituent pourtant des opérations ou capacités dont la maîtrise est recherchée à l'occasion de l'action de formation. On comprend pourquoi la pédagogie par objectifs laisse ouvert le problème des apprentissages. Elle travaille en deçà et ne dit rien du cheminement qui permet de construire la compétence visée. A fortiori, lorsque l'objectif se réfère à un objet "absolu" : un savoir à acquérir, ou une opération intellectuelle à maîtriser, ni

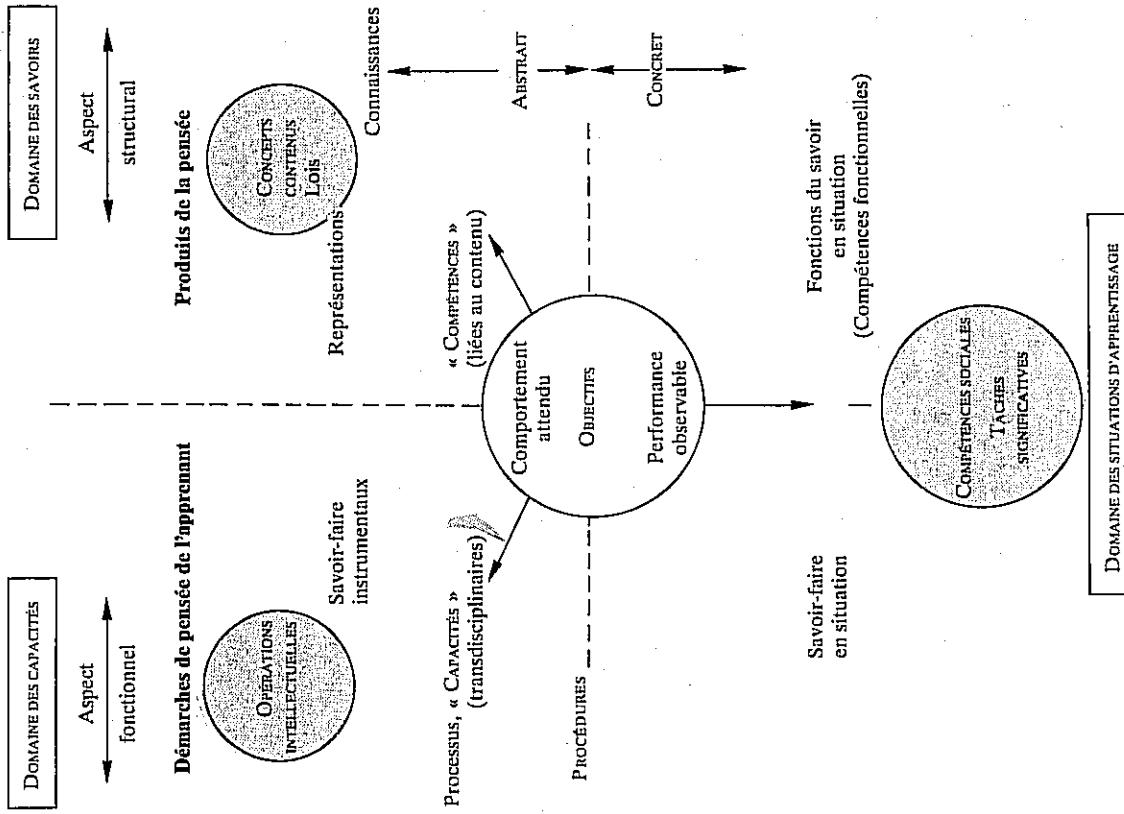


Fig. 6. – Les objets possibles des objectifs pédagogiques

¹⁶J. CARDINET, *Evaluation scolaire et pratique*, op. cit., p. 218.
¹⁷Ibid., p. 36.

L'évolution de la structure cognitive, ni le développement des fonctions intellectuelles, ne sont donnés à l'observation. La difficulté pour l'évaluation formative est donc de se doter de critères qui, au-delà de la performance cible, scandent bien une évolution positive des structures et fonctions de l'apprenant. Or celles-ci ne sont pas observables directement, mais seulement à travers leurs produits. Pour mettre en œuvre efficacement une démarche d'évaluation formative, il faudrait, et nous y reviendrons, posséder un modèle de la progression cognitive, et un tableau de correspondance performance/compétence permettant d'apprécier l'évolution de la compétence à travers l'amélioration de la performance.

C'est pourquoi certains ont proposé de se centrer résolument sur la tâche et de mettre l'évaluation au service, d'abord, d'une amélioration des produits, celle-ci étant observable et appréciable par rapport à des critères susceptibles d'être objectivement définis. Telle est l'idée de départ de l'*évaluation formatrice*. Considérant que le travail de l'élève est de parvenir à la réussite dans certaines tâches déterminées (ex. : faire une dissertation, écrire un texte narratif, résumer un texte écrit), des enseignants-rechercheurs de l'Université de Provence 18 se sont interrogés sur les conditions de possibilité de cette réussite et sur la façon dont l'évaluation pouvait y contribuer. La prise en compte de l'importance de la représentation des buts à atteindre, et la conviction que seul l'apprenant est vraiment capable de réguler son activité d'apprentissage, car seul capable de connaître ses processus et de les corriger, les ont conduits à affirmer que l'évaluation doit viser avant tout l'appropriation par l'apprenant de deux séries de critères, dont la connaissance et la possession sont de nature à favoriser la réussite dans la tâche, c'est-à-dire la réalisation d'un "bon" produit.

- *les critères ou signes de réussite*, qui permettent de juger le produit fini ;
- *les critères de réalisation*, qui correspondent aux exigences objectives de la tâche, autrement dit à des opérations qu'il faut absolument effectuer pour la réaliser.

Par exemple, pour un résumé de texte, constitueront des critères de réalisation :

- la reprise des idées essentielles, et d'elles seules ;
- le respect des proportions ;
- l'utilisation des termes de liaison.

Les critères de réussite préciseront, pour chaque critère de réalisation, le seuil d'acceptabilité. Par exemple : pas d'oubli d'idée essentielle ; présence de termes clés, etc.

Ces différents critères pourront être trouvés par les apprenants eux-mêmes, à l'occasion d'une analyse de produits déjà réalisés. Ce qui fait de

l'évaluation formatrice une pratique résolument pédagogique, centrée sur les démarches des élèves, et privilégiant l'autorégulation.

Cette démarche a de nombreux mérites. Elle conduit à confronter l'élève à une tâche qui, dans sa relative complexité, a plus de chance d'être mobilisatrice qu'un objectif qui, dans la simplicité de son énoncé opérationnel, reste formel et artificiel. Elle permet d'éviter le piège de la parcelisation en micro-objectifs. Elle priviliege ce que l'élève construit, par rapport à ce qu'on lui donne. Elle rend inutile finalement la possession d'un modèle précis de la compétence cognitive ou du développement intellectuel, comme celle d'un modèle de fonctionnement des savoirs à enseigner (nature du contenu), ce qui est une façon élégante de surmonter l'écueil de la limitation actuelle des connaissances dans le double domaine de l'analyse didactique des savoirs et de l'analyse psychologique des compétences.

Toutefois, ces avantages ont leurs revers, et l'évaluation formatrice appelle deux remarques importantes. Tout d'abord, est-il encore question d'évaluation ? Le souci dominant – et louable – est d'accompagner et d'aider l'apprentissage. De ce point de vue, l'évaluation formatrice est une forme achevée d'évaluation formative. Mais la recherche de l'efficacité pédagogique fait pratiquement disparaître l'évaluation au sens propre, qui n'intervient plus que comme évaluation d'un produit fini. Paradoxalement, et c'est-à-dire la réalisation hors de portée, on se contente d'évaluer les résultats. De ce point de vue, l'évaluation formatrice est le symétrique de la pédagogie par objectifs (fig. 7). Dans l'acte d'apprentissage, le

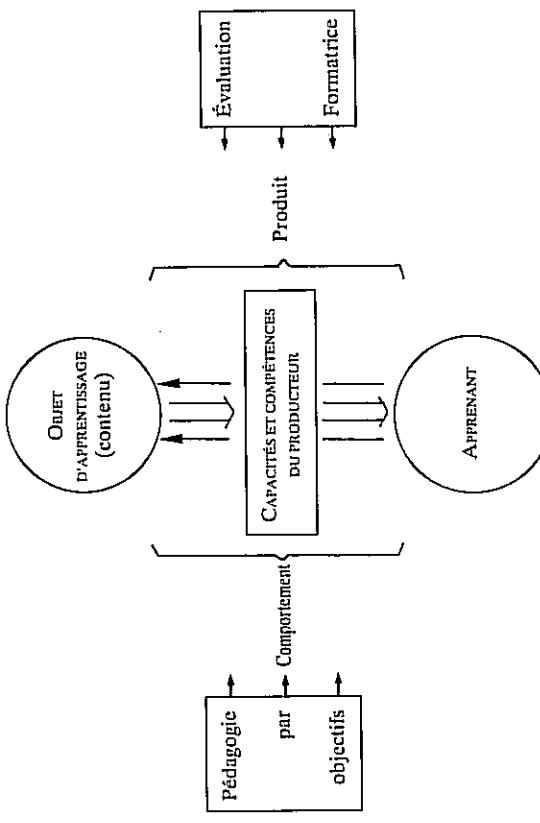


FIG. 7. — Pédagogie par objectifs et évaluation formatrice