

## **Cinquième partie**

### **L'évaluation**

# Chapitre 17

## Généralités

17.1	Un vieux problème ... jamais résolu . . . . .	577
17.2	Que veut-on évaluer ? . . . . .	578
17.2.1	Évaluer des comportements . . . . .	579
17.2.2	Un problème de stabilité . . . . .	581
17.3	Que veut-on faire des résultats de l'évaluation ? . . . . .	584
17.3.1	L'évaluation formative . . . . .	585
17.3.2	L'évaluation certificative . . . . .	586
17.4	Comment évaluer ? . . . . .	588
17.4.1	Choisir des questions . . . . .	589
17.4.2	Noter les prestations des élèves . . . . .	590
17.5	Le problème de l'évaluation des résolutions de problèmes . . . . .	593
17.6	Résumé des chapitres suivants . . . . .	595

## 17.1. Un vieux problème . . . jamais résolu

La littérature consacrée à l'évaluation remplirait des bibliothèques. Il serait cependant hasardeux d'affirmer qu'il y ait aujourd'hui accord sur les principes et surtout sur les méthodes de l'évaluation, même si d'importants efforts de clarification ont incontestablement été fournis par de nombreux auteurs depuis 20 ou 30 ans. Invités — souvent de façon prématurée — à modifier leurs pratiques en appliquant des réformes insuffisamment expérimentées, les enseignants ont pour l'essentiel conservé leurs habitudes derrière la façade d'une présentation modernisée des « bulletins ».

Faute de la compétence nécessaire et d'une connaissance approfondie de la littérature, il ne nous est pas possible ici de dresser l'état de la question. Nous nous contenterons — sans prétendre à aucune originalité — de formuler un certain nombre d'observations naïves, liées — autant que possible — au nouveau (?) point de vue « *apprentissage de compétences* ». Bien entendu, nous nous centrerons sur le cas des mathématiques, ce qui n'empêche que certaines de nos remarques pourraient s'appliquer à d'autres disciplines.

Dans tout acte d'évaluation, plusieurs questions — non indépendantes entre elles — doivent être abordées.

- Que veut-on évaluer ?
- Que veut-on faire des résultats de l'évaluation ?
- Comment évaluer ?

Il peut paraître étonnant que nous considérions la question « Comment évaluer ? » en dernier lieu, à la suite de « Que veut-on faire des résultats de l'évaluation ? ». C'est que la façon d'évaluer dépend précisément de ce que l'on veut faire des résultats : on n'évalue pas de la même manière dans le cas d'une évaluation formative que dans le cas d'une évaluation certificative.

## 17.2. Que veut-on évaluer ?

Dans un enseignement qui se fixe comme objectif l'apprentissage de compétences par les élèves, il semble assez naturel que l'évaluation porte sur l'acquisition de ces compétences. Il est donc important de souligner en premier lieu que

Ce ne sont pas les élèves que l'on évalue,  
mais l'efficacité des compétences qu'ils ont acquises.

Évaluer les élèves serait porter un jugement, ou tout au moins une appréciation, sur des individus en tant que tels, des individus qui peuvent encore évoluer fortement. Mais alors, qu'évaluons-nous ?

### 17.2.1 Evaluer des comportements

On ne voit pas comment il serait possible d'évaluer l'acquisition d'une compétence par un élève sans demander à celui-ci d'avoir certains comportements observables en réponse à certaines questions.

On ne peut évaluer que des comportements observables.

Ce principe *behaviouriste* ne doit pas nous amener à accorder une place exclusive à l'apprentissage de comportements. Dès lors qu'une compétence est *l'aptitude à mettre en œuvre un ensemble organisé de savoirs, de savoir-faire et d'attitudes permettant d'accomplir un certain nombre de tâches*, et qu'un concept rassemble des savoirs et des procédures, nous pouvons considérer qu'une compétence est basée sur la mise en œuvre d'un ou plusieurs concepts.

Plusieurs remarques s'imposent :

1. Comme un animal, un enfant peut être dressé à reproduire certains comportements de façon plus ou moins efficace. L'observation de l'utilisation correcte de procédures ne permet donc pas de conclure automatiquement à l'acquisition des concepts sous-jacents.
2. Pour permettre à l'élève d'avoir des comportements efficaces, un concept ne doit que rarement avoir atteint une forme « définitive » (pour autant qu'une telle chose soit possible). Certains modèles mentaux intermédiaires suffisent pour résoudre un certain nombre de problèmes.

EXEMPLE 17.2.1 *Il n'est pas nécessaire d'avoir exclu le nombre 1 de la liste des nombres premiers pour effectuer les décompositions en facteurs premiers utiles dans le calcul des fractions.*

3. Si un concept apparaît insuffisamment formé pour que l'élève puisse faire face aux problèmes qui lui sont soumis, l'évaluation doit s'efforcer de déterminer quel stade a été atteint. En particulier, quel modèle mental l'élève s'est-il forgé ? Les comportements observés, les procédures utilisées, doivent permettre de répondre à cette question.

4. Un concept peut être mis en œuvre à travers des procédures différentes mais équivalentes en ce qu'elles fournissent les mêmes résultats à partir des mêmes données. On connaît par exemple plusieurs méthodes de multiplication écrite. L'évaluation doit donc être organisée de manière à permettre à l'élève d'appliquer éventuellement une procédure que l'examineur n'a pas prévue.

Toute évaluation porte sur des comportements et présente de ce fait un caractère temporaire, c'est une « photographie instantanée » comme on disait autrefois. Mais cette photographie n'est-elle pas parfois floue ? Il est bien connu qu'un élève peut en fonction du contexte avoir des réactions différentes à une même sollicitation.

### 17.2.2 Un problème de stabilité

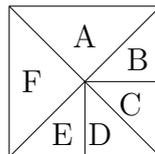
Différentes études de didactique expérimentale des mathématiques (voir notamment [90], [121]) ont mis en évidence, entre autres, les phénomènes suivants :

- Il suffit parfois de permuter deux questions d'un questionnaire pour que les taux de réussite à l'une (au moins) de ces deux questions soient fortement modifiés.
- De façon analogue, si un enseignant juge une question difficile, il sait très bien qu'en la faisant précéder d'une question préparatoire il est souvent possible d'aiguiller l'élève vers la solution, alors même que cette question préparatoire n'est pas nécessairement en rapport direct avec la question principale.
- Une question portant sur un sujet mal maîtrisé par un élève peut induire chez celui-ci une régression l'amenant à commettre une erreur que l'on aurait pu croire disparue.
- Si en répondant à un questionnaire, un élève arrive lors de la résolution d'une question à un résultat en contradiction avec celui obtenu à une question précédente, ou avec un fait mathématique connu, un « retour en arrière » est parfois observé qui amène à la correction d'une erreur qui — sinon — aurait subsisté.

Pour illustrer certains des comportements qui viennent d'être décrits, contentons-nous de relater une expérience datant de 1980 (voir [105]) :

**EXEMPLE 17.2.2** *Lors d'une étude sur l'intuition en probabilité, une question de calcul avait été posée à un échantillon d'environ 180 élèves de 6<sup>e</sup> année de l'enseignement secondaire général. La question admettait les deux variantes indiquées ci-dessous, chaque élève n'ayant à répondre qu'à l'une d'entre elles.*

- *Variante A : Je lance sans viser un pion trois fois de suite sur ce carré :*



*Quelle est la probabilité de faire apparaître la suite ABF ?*

- Variante B : La situation est la même, mais le pion est lancé 6 fois et on demande la probabilité de la suite ABCDEF.

Les taux de réussite à ces deux variantes ne sont pas très différents : 39% pour la variante A, 34 % pour la variante B. Les deux variantes semblent en effet — à première vue — du même niveau de difficulté : si la seconde nécessite le calcul d'un produit de 6 facteurs alors que la première n'en fait intervenir que 3, on peut penser qu'il n'y a là, pour des élèves de dernière année du secondaire, rien qui justifie une différence de réussite significative. Tout au plus doit-on s'attendre à rencontrer un plus grand nombre de mauvaises réponses dues à des erreurs de calcul. C'est apparemment ce qui se passe.

Une étude plus fine des réponses, portant sur les comportements erronés qui sont apparus, met cependant en évidence des phénomènes intéressants. On peut ainsi se demander si les élèves ont appliqué une règle d'addition ou une règle de multiplication. Dans la variante A, les élèves ayant appliqué une règle d'addition sont repérables à la réponse fournie :  $\frac{5}{8}$ . Cette réponse est le fait de 32% des élèves.

Mais dans la variante B, ce pourcentage n'est que de 8%. Nous expliquons cela par le fait que le comportement d'addition fournit cette fois 1 comme résultat, et que les élèves connaissaient le fait que 1 est la probabilité de l'événement certain. La plupart des élèves qui ont réalisé une addition ont alors abandonné cette méthode et se sont tournés vers autre chose.

Certains ont réalisé une multiplication : le taux d'apparition de ce comportement (y compris les bonnes réponses) passe de 47% pour la variante A à 56% pour la variante B. Mais surtout, beaucoup ont opté pour une équiprobabilité en répondant  $\frac{1}{2}$ . Ce comportement est le fait seulement de 6% dans la variante A, mais de 34% dans la variante B.

Dans cet exemple, nous voyons intervenir le phénomène de retour en arrière après la constatation d'une contradiction avec un fait connu et celui de régression par réapparition d'une erreur plus grossière.

Il faut aussi tenir compte de ce que le stress de l'examen ou même des circonstances éventuellement totalement étrangères au cours de mathématique ou à l'école, sont susceptibles de modifier profondément le comportement d'un élève.

Nous résumerions volontiers ces constatations en disant que

le comportement d'un élève apparaît souvent comme une fonction discontinue d'un certain nombre de paramètres qui ne nous sont pas tous connus et *a fortiori* que nous ne contrôlons pas tous. (A moins qu'il ne s'agisse de « chaos déterministe », ce qui revient à peu près au même.)

Nous constatons donc que

nous devons évaluer la stabilité des comportements.

Bien entendu, nous devons d'abord nous demander quel degré de stabilité nous exigeons pour déterminer si une compétence est acquise. L'exemple classique du pilote d'avion de ligne nous fait prendre conscience de ce que la stabilité d'un comportement doit dans certains cas être absolue : pour recevoir son brevet, le candidat-pilote ne peut se contenter de réussir 95% de ses décollages et de ses atterrissages !

Des trois phases distinguées par A. Sfard lors de la formation d'un concept (et des compétences qui lui sont associées), c'est au cours de la phase de réification que la stabilité peut être atteinte. Par contre, au cours des phases précédentes, des images mentales différentes se succèdent, s'affinent, le concept est en formation, les comportements sont instables.

Pour vérifier l'acquisition d'une compétence, c'est-à-dire la stabilité des comportements associés, nous devons

- soit répéter l'évaluation en faisant varier le contexte,
- soit vérifier autrement que le stade de réification a été atteint, donc que le ou les nouveaux concepts sont effectivement devenus des objets auxquels de nouvelles procédures peuvent s'appliquer.

La seconde possibilité suppose que les différentes phases (caractérisées notamment par des comportements d'erreur spécifiques) qui marquent l'apprentissage de chaque nouveau concept mathématique soient bien connues, ce qui est loin d'être toujours le cas.

Par ailleurs, si pour être pleinement conceptualisée une notion nouvelle nécessite que de nouvelles procédures — d'un niveau plus élaboré — lui soient appliquées, amorçant ainsi le début d'un autre processus de conceptualisation, l'idée même de *compétence* TERMINALE devient sujette à caution. De façon plus précise, l'enseignement devrait être poussé au-delà des compétences déclarées terminales par les textes officiels sans quoi leur *certification* ne pourrait être que douteuse. C'est parfois — faut-il le dire ? — ce que montrent les résultats obtenus par les étudiants lorsqu'ils abordent l'enseignement supérieur. Mais cette remarque nous amène au paragraphe suivant.

### 17.3. Que veut-on faire des résultats de l'évaluation ?

La distinction entre *évaluation formative* et *évaluation certificative* est bien connue. Nous ne la rappellerons donc pas en détails ici. Dans ce qui précède, nous n'avons pas distingué les deux aspects de l'évaluation. Ils sont au surplus parfois difficiles à distinguer.

Lorsque nous demandons qu'à travers les comportements erronés d'un élève on détecte quelle phase de conceptualisation il a atteinte sur un sujet donné, nous faisons incontestablement de l'évaluation formative puisque nous devrions pouvoir en déduire un traitement de remédiation approprié. Mais si on parvient à surmonter d'évidentes difficultés matérielles, pourquoi un diplôme ne pourrait-il pas rassembler de telles informations en ce que nous appellerions le « profil cognitif » d'un individu ? A l'employeur de décider alors si ce profil lui convient. Il s'agirait bien dans ce cas d'une évaluation certificative. A l'heure actuelle, ce profil cognitif se réduit à une note <sup>(1)</sup>, un pourcentage. Quel usage un employeur peut-il faire de ce renseignement imprécis ?

---

<sup>(1)</sup> Dans ce travail, nous utiliserons indifféremment les expressions *note*, *notation* utilisées en France et les expressions *cote*, *cotation* utilisées dans le même sens en Belgique.

### 17.3.1 L'évaluation formative

Elle a toujours été pratiquée, consciemment ou non, formellement ou non. En fait chaque fois qu'un enseignant détecte chez un élève l'apparition d'un comportement erroné et prend les mesures adéquates pour le rectifier, il pose un acte d'évaluation formative. Il n'est pas nécessaire pour cela que l'évaluation ait été quantifiée, c'est-à-dire que l'élève ait fait l'objet d'une note numérique.

La pratique de la note lors d'une évaluation formative, en dramatisant la situation, en culpabilisant éventuellement l'élève, aurait plutôt l'effet inverse de celui qui est recherché. Le but est en effet d'abord *d'exploiter l'erreur d'un élève en vue de le faire progresser dans un processus de conceptualisation*. Par exemple, on peut être amené à lui montrer qu'il s'est formé une image mentale qui lui permet certes de résoudre de façon correcte un certain nombre de problèmes, mais qui s'avère insuffisante et l'amène à des erreurs dans le cas d'autres situations problématiques.

**EXEMPLE 17.3.1** *Déterminer les asymptotes éventuelles de la fonction  $x \mapsto |x|$ .*

La plupart des élèves de 6<sup>e</sup> répondront à cette question que la fonction donnée n'admet pas d'asymptote. Leur image mentale d'asymptote est celle de « droite ayant en commun avec une courbe un et un seul point situé à l'infini ». Quelques-uns — peut-être plus accrochés à des procédures — rechercheront des asymptotes obliques en calculant  $\lim_{x \rightarrow \pm\infty} \frac{|x|}{x}$  et trouveront les deux droites  $y = x$  et  $y = -x$ . Un résultat analogue peut être obtenu pour la fonction  $\frac{\sin x}{x}$ .

La mise en évidence de cette contradiction et la discussion qui en résulte permettent de dépasser l'image mentale primitive de « droite tangente à l'infini » et de la remplacer par celle de « meilleure approximation de la courbe par une droite ». Quelle droite pourrait le mieux approcher une droite sinon cette droite elle-même ?

C'est en fait un concept d'approximation qui commence ainsi à se former.

L'évaluation formative n'a de signification que durant un apprentissage. Elle doit être vécue par l'élève comme une aide qui lui est fournie.

### 17.3.2 L'évaluation certificative

Elle a deux aspects :

1. du point de vue de la société, elle a pour but de *sélectionner* les individus en déterminant *quelle place active* ils sont capables de prendre dans la vie économique, sociale et culturelle.
2. du point de vue de l'élève, elle entraîne une *valorisation* ou une *dévalorisation*.

Ces deux aspects soulèvent des problèmes délicats, de nature socio-politiques qui dépassent largement le point de vue technique que nous adoptons ici. Nous nous contenterons de remarquer que les documents relatifs aux *socles de compétences* et aux *compétences terminales* fixent les contenus sur lesquels peut porter la certification, sans préciser de quelle façon celle-ci s'effectue.

Plusieurs questions se posent :

1. Faut-il évaluer les compétences dites *transversales*. Dans [30], J. Beckers estime *inopportun et même dangereux de prévoir une évaluation de ces compétences pour elles-mêmes, en dehors de tout contenu*. En fait, à l'heure actuelle, personne ne voit clairement comment on pourrait procéder à une telle évaluation.
2. Quel est le seuil de réussite à atteindre afin d'obtenir la certification ? En d'autres termes, faut-il que les socles ou les compétences terminales soient atteints à 100%, 95%, 50% ou tout autre pourcentage qui serait fixé pour toutes les écoles ?

Si les socles et les compétences terminales sont des contenus *minimaux*, dont la fixation est destinée à assurer l'homogénéité de la formation des élèves, il serait logique d'exiger qu'ils soient réussis à 100%. Dans ce cas il n'y aurait plus que deux catégories d'élèves : ceux qui échouent et ceux qui réussissent.

On peut se demander si la certification fournirait alors les informations suffisantes permettant aux élèves de choisir valablement une orientation d'études dans l'enseignement supérieur. À moins que l'on considère que les examens puissent comporter deux parties : une serait limitée aux contenus minimaux et assurerait la certification. L'autre pourrait déborder des contenus minimaux et permettrait aux élèves de donner le meilleur d'eux-mêmes dans les branches de leur choix.

Si les socles et les compétences terminales ne sont pas des contenus *minimaux*, on ne distingue plus clairement quelle est leur raison d'être.

3. Comment évaluer des compétences globales telles que la résolution de problèmes ? Nous reviendrons sur cette question dans la suite, mais d'ores et déjà nous pouvons remarquer qu'une compétence globale N'EST PAS la somme des compétences particulières qu'elle met en œuvre. Pour réussir une épreuve destinée à évaluer une compétence globale, un élève doit non seulement maîtriser les compétences particulières, mais il doit aussi être capable de déterminer celles qui doivent intervenir et de les coordonner. La forme traditionnelle d'examen se prête mal à des tâches de cette nature.

## 17.4. Comment évaluer ?

Ici aussi, nous distinguerons l'évaluation formative et l'évaluation certificative. Pour la première, *tous les moyens sont permis*. Selon la formule de J. Beckers ([30]), ils ne relèvent que de la responsabilité de l'enseignant.

Encore peut-on attirer l'attention sur quelques points. Nous avons déjà remarqué que la note chiffrée est inopportune dans ce cas. Ce qui est plus important est que l'enseignant connaisse bien les types d'erreurs que ses élèves sont susceptibles de commettre, de façon à les repérer et les exploiter rapidement. Parmi les outils pédagogiques à mettre à la disposition des enseignants devraient figurer des compte-rendus des travaux des chercheurs qui étudient les processus d'apprentissage et ont obtenu dans le cas de l'enseignement des mathématiques des résultats non négligeables depuis une vingtaine d'années.

L'enseignant doit aussi bien connaître quels sont les objectifs qu'il veut atteindre et qui ne sont pas nécessairement ceux des socles ou des compétences terminales, parfois trop imprécis.

Quant à la technique d'évaluation certificative, nous pourrions la résumer comme suit :

1. Etablir une liste de questions représentatives des contenus à évaluer et des niveaux visés.
2. Soumettre ces questions aux élèves, oralement ou par écrit, en classe ou à domicile.
3. Noter les prestations des élèves.

Mais tout cela ne va pas sans de grosses difficultés.

### 17.4.1 Choisir des questions

Comment s'assurer que les questions choisies sont adéquates ? Il est prévu d'établir des banques d'exercices standardisés pouvant servir de référence. Il sera nécessaire que ces exercices standardisés soient étalonnés sur une population d'élèves suffisamment importante. Mais même si les enseignants ont à leur disposition une telle banque d'exercices, les problèmes ne sont pas nécessairement résolus.

En rédigeant un exercice sur un modèle figurant dans la banque, on peut facilement sans y prendre garde apporter une modification même légère dont les conséquences sont importantes. Par ailleurs, un ensemble d'exercices standardisés ne constitue pas un test standardisé. Comme nous l'avons déjà signalé, les interactions entre les exercices d'un même questionnaire peuvent modifier fondamentalement les comportements des élèves. N'oublions pas que ceux-ci *apparaissent souvent comme des fonctions discontinues des paramètres de la situation, notamment du contexte.*

Pas plus que les examens traditionnels, les exercices standardisés ne testeront pas les élèves, ni leurs compétences mais uniquement des comportements ponctuels. Le problème de la stabilité de ces comportements demeure.

## 17.4.2 Noter les prestations des élèves

### 17.4.2.1 Des notes numériques ?

La notation traditionnelle repose sur l'attribution d'une note numérique (sur 10, 20 ou 100, peu importe).

Dans [77], Georges GLAESER fait un procès convaincant de cette pratique à laquelle il reproche six défauts majeurs.

1. Il n'y a aucune raison pour que les performances d'un élève lors d'un test puissent être représentées valablement par une note numérique, c'est-à-dire dans un espace de dimension 1. En dehors de cas marginaux, ce sont des notes de nature vectorielle qu'il faudrait utiliser pour rendre compte des différentes facettes des comportements d'un élève.
2. Les notes sont trop précises. On sait depuis les premières études de docimologie que des correcteurs différents peuvent attribuer à une même copie des notes non seulement différentes, mais même dont les écarts sont plusieurs fois supérieurs à l'intervalle qui sépare deux notes consécutives sur l'échelle utilisée.
3. Les notes ne devraient pas constituer un ensemble totalement ordonné. Tout simplement parce que ce n'est pas le cas non plus des élèves.
4. Les calculs effectués sur des notes n'ont aucune signification. Ajouter la note obtenue en histoire à celle obtenue en mathématique est aussi absurde que d'ajouter des pommes et des poires. Il en est de même au sein des mathématiques : il n'y a de sens à additionner les notes attribuées à deux questions que si celles-ci testent la même capacité. Si par contre, elles testent des aptitudes non corrélées, l'addition (ou la moyenne) n'a aucun sens.
5. L'échelle numérique n'est pas nécessairement linéaire : il peut y avoir un écart plus grand entre deux performances notées 7/10 et 8/10 qu'entre deux performances notées 5/10 et 6/10.
6. La barre de sélection est placée arbitrairement. Faut-il faire échouer les élèves ayant moins de 4/10 ? de 5/10 ? de 6/10 ?

Une des meilleures preuves du caractère arbitraire de l'échelle numérique est constituée par les manipulations auxquelles se livrent parfois des enseignants pour avoir parmi leurs élèves un peu d'échecs (sinon on dira qu'ils jettent les points à la tête des élèves), mais pas trop (sinon, on dira que ce sont des vaches).

Non seulement on peut dire que la notation numérique est inadaptée, mais aussi qu'il s'est instauré un « culte » de la courbe de Gauss.

D'après J.-C. Parisot (voir [48]), le modèle gaussien de distribution des notes ne peut être considéré comme une référence d'exactitude lors d'une évaluation certificative que lorsque l'examen a été préalablement testé et petit à petit mis au point sur de grandes populations, et qu'un protocole de correction est fourni.

Bref, seule une épreuve de type *baccalauréat* peut accepter une distribution gaussienne des résultats. Toute autre épreuve, plus ponctuelle et à moindre échelle, n'admet que rarement une telle distribution. En effet, pourquoi devrait-on obtenir une distribution représentative d'un comportement aléatoire à l'issue d'une séquence d'enseignement ou justement tout a été fait pour contrer le hasard ?

### 17.4.2.2 Et sinon quoi ?

Il convient évidemment de se demander ce qui pourrait remplacer la notation traditionnelle. Tant qu'il s'agit d'évaluer l'acquisition de compétences particulières précises, facilement observables et d'ampleur limitée, la liste de celles qui sont atteintes par un élève peut être établie. On aboutit ainsi à une situation dans laquelle la note d'un élève n'est pas une grandeur scalaire mais une grandeur vectorielle.

Mais on sait bien qu'il n'est pas possible d'en rester là. Des activités de nature plus globale doivent être soumises aux élèves et évaluées faute de quoi l'enseignement des mathématiques risque de perdre toute signification.

Il convient alors de parler de l'analyse de données, qui sous diverses formes permet parfois dans le domaine de l'évaluation le contrôle des erreurs (d'évaluation) et l'instauration d'une échelle de mesure.

Il n'est pas possible ici de faire une liste commentée des diverses méthodes employées, ni de donner une théorie générale. Nous utiliserons des méthodes de ce type aux chapitres 19 et 21. Mais dès à présent, il nous semble important de soulever, plusieurs inconvénients de ces méthodes :

- pour atteindre une certaine objectivité, et éviter le plus possible d'obtenir des résultats ne correspondant à rien, il est nécessaire d'employer conjointement plusieurs types d'analyse ;
- ces analyses, même si elles peuvent se contenter de porter sur une faible population, ne permettent des conclusions *généralisables* que lorsque la population testée est vaste ;
- la maîtrise de ces méthodes d'analyse et le temps qu'elles exigent suppose que des professeurs se spécialisent dans ce domaine et s'écartent de leur tâche purement pédagogique (ainsi que de la résolution de problèmes !), ce qui peut mener à une dichotomie du corps enseignant (avec les difficultés de coordination que l'on peut imaginer).

Comme on le constate, aucune des méthodes envisagées jusqu'ici ne garantit une évaluation objective. Peut-être faut-il se demander si la recherche d'une telle objectivité n'est pas une illusion !

Il convient en tous cas de remarquer que tout acte d'évaluation consiste à comparer les comportements d'un élève à des comportements modèles, qui sont le plus souvent ceux qu'aurait l'examineur s'il était placé dans la situation de l'élève. Il n'est pas nécessaire d'en dire plus pour établir la difficulté d'élaborer une méthode d'évaluation objective. Même des questions standardisées ne résoudraient pas le problème. A moins qu'elles soient accompagnées d'une grille de correction qui imposerait à tous les correcteurs la subjectivité de l'auteur de la grille. L'objectivité n'est-elle qu'une subjectivité imposée ?

## 17.5. Le problème de l'évaluation des résolutions de problèmes

Dans la pratique des classes, peu de temps est réservé à la résolution de problèmes. On peut trouver à cela des raisons diverses, notamment :

- Les enseignants ne sont guère formés à ce type d'activité et encore moins à leur évaluation.
- Une activité de résolution de problèmes *isolée dans le temps* est peu porteuse d'informations, tant pour les élèves que pour les professeurs : les élèves peu habitués à résoudre des problèmes éprouvent des difficultés à troquer une attitude passive basée sur l'attente de ce qui va arriver pour une attitude active de recherche. Ces difficultés masquent les avantages qui pourraient résulter de l'activité proposée. L'évaluation de l'activité est alors sans intérêt.  
On pourrait comparer la résolution de problèmes à un sport : l'entraînement, le progrès, la reconnaissance ne prennent leur sens que dans la durée.
- L'école n'est pas isolée dans la société. Les mutations de celles-ci ont des répercussions sur les activités et attitudes des élèves et des enseignants. Il en résulte une instabilité qui handicape d'autant plus le travail des enseignants.

Ces raisons ne doivent pas empêcher la réflexion et l'action. Nous pouvons formuler quelques hypothèses de travail et en tirer des conséquences

1. La résolution de problèmes nécessite des méthodes diverses auxquelles les élèves doivent être initiés. Évaluer la résolution de problèmes nécessite que cette initiation soit conçue comme un véritable enseignement. Ainsi, malgré les difficultés,

il faut construire un enseignement de la résolution de problèmes

2. L'élève a le droit de se tromper, de ne pas arriver au bout du problème. Même dans ce cas, l'activité de recherche est profitable.

L'évaluation doit porter sur la démarche de recherche telle qu'elle a été décrite dans les chapitres 6 à 8, ainsi que sur la démonstration, en tant qu'apprentissage de la méthode critique, plus que sur l'obtention de la solution du problème posé.

3. Tout apprentissage est relatif. Un problème n'a que rarement une seule bonne solution. L'activité de résolution de problèmes, les stratégies utilisées, l'utilisation des ressources, la communication, ont aussi un aspect relatif et non absolu.

L'évaluation doit tenir compte du côté relatif de l'activité.

4. Le droit pour l'élève à se tromper, le côté relatif de l'activité de résolution, ne permettent pas d'utiliser des résolutions de problèmes dans le cadre d'une évaluation exclusivement certificative.

L'évaluation de la résolution d'un problème est d'abord formative. Moyennant une réforme de la méthode de notation, elle pourrait être également certificative.

C'est donc une compétence à développer plutôt qu'une compétence à atteindre.

5. Enfin, si on veut agir sur la vie quotidienne des classes, il importe que

toute procédure d'évaluation de la résolution d'un problème soit fondamentale et simple.

## 17.6. Résumé des chapitres suivants

Dans les chapitres suivants, nous présenterons plusieurs expériences d'évaluation réalisées soit dans des classes de l'enseignement secondaire, soit à l'université.

Vu la place que nous accordons à la résolution de problèmes et à la problématisation d'un cours de mathématique, il n'était que normal que nous nous attachions à dégager des pistes en vue de faciliter l'évaluation des élèves placés dans de telles situations. Cette réflexion est rapportée aux chapitres 18 et 19.

Au chapitre 20, on envisage le cas d'élèves confrontés à un autre type d'activités : la réalisation d'un dossier-projet. Dans ce cas, l'élève doit faire preuve non seulement de ses qualités scientifiques, mais aussi de capacités de rédaction et de communication qui figurent également parmi les compétences terminales importantes.

Enfin, au chapitre 21, nous présentons une évaluation basée sur un questionnaire plus étendu que ceux qui ont été traités aux chapitres 18 et 19. Il s'agit là d'un examen universitaire ayant des spécificités propres mais qui a l'avantage de mettre en évidence un comportement peut-être moins fréquent dans l'enseignement secondaire : la non-réponse à une question. Dans quelle condition peut-on mettre l'absence de réponse sur le même pied qu'une erreur ?

Ces tentatives d'évaluation sont menées selon des méthodes différentes : aux chapitres 18 et 20, seules des analyses approfondies des copies des élèves sont pratiquées alors qu'aux chapitres 19 et 21, des méthodes d'analyse statistique sont mises en œuvre.

Cependant, ces méthodes ont des caractéristiques communes. La plus visible est la quantité de travail importante qu'elles fournissent à l'enseignant. Il ne nous appartient pas de tirer argument de ce fait dans un sens ou dans l'autre mais les décideurs ne pourront le négliger. Plus importante — du point de vue qui nous préoccupe, celui de l'évaluation des élèves — est la place que les différentes méthodes accordent à la détermination des *comportements* qu'ont eu les élèves durant l'évaluation. Par comportement, nous entendons les démarches, méthodes, algorithmes, . . . , qu'ils ont (ou n'ont pas) mis en œuvre en vue d'atteindre l'objectif qui leur était fixé.

La détermination des comportements des élèves est le point crucial tant de l'évaluation formative que de l'évaluation certificative.

A travers l'étude des comportements des élèves apparaît aussi la possibilité, et même la nécessité, d'évaluer le questionnaire soumis aux élèves lui-même. La question principale est de savoir si ce questionnaire est en harmonie avec l'enseignement qui a été dispensé : si chacun s'accorde à dire qu'on ne peut évaluer que ce qui a été enseigné, il n'est peut-être pas rare que des questions d'examen ne se situent pas au même niveau que l'enseignement. Nous en verrons un exemple au chapitre 21.

Une autre caractéristique commune aux évaluations faites (ou esquissées) est le fait qu'elles conduisent à une notation vectorielle. Comme il a été dit précédemment, ce n'est que logique puisque les comportements qui sont évalués peuvent n'être aucunement corrélés. Toutefois, au chapitre 20, nous utilisons une méthode *d'aide à la décision* en vue de construire une note *numérique* qui GLOBALISE la note vectorielle obtenue. Les décisions à prendre ne s'accommodent guère du vectoriel. . .

Les méthodes statistiques utilisées aux chapitres 19 et 21 sont lourdes et nécessitent des logiciels spécialisés et surtout de bonnes connaissances statistiques pour qui veut les appliquer en comprenant ce qu'il fait. Leur manque de transparence en rend les conclusions difficiles à expliquer (aux parents par exemple). Leur avantage est de diminuer dans une certaine mesure la subjectivité de l'évaluation.

Cette subjectivité ne saurait disparaître entièrement : ce sera toujours à l'enseignant de déterminer quels comportements il décide de recenser chez ses élèves. Mais cette sélection étant effectuée, il lui « suffira » de noter la présence ou l'absence de ces comportements chez chaque élève : une évaluation binaire en 0 ou 1. Les logiciels statistiques prennent alors le relais pour calculer des notes (vectorielles ou scalaires) plus nuancées. Les résultats semblent intéressants mais mériteraient d'être confirmés et affinés par de nouvelles études.

Pour clore ce chapitre, répétons ce que nous avons dit dans l'introduction : les recherches sur l'évaluation doivent se poursuivre.

### Références

[90], [121], [105], [30], [77], [48].