

Chapitre 1

L'évaluation
décortiquée**PETIT TEST**

Pour chacune des notions suivantes, indiquez par une croix si vous pouvez en donner une définition pertinente¹.

	OUI	NON
1 Compétence	<input type="checkbox"/>	<input type="checkbox"/>
2 Capacité	<input type="checkbox"/>	<input type="checkbox"/>
3 Niveau de compétence	<input type="checkbox"/>	<input type="checkbox"/>
4 Référentiel	<input type="checkbox"/>	<input type="checkbox"/>
5 Progression	<input type="checkbox"/>	<input type="checkbox"/>
6 Savoirs	<input type="checkbox"/>	<input type="checkbox"/>
7 Savoir-faire	<input type="checkbox"/>	<input type="checkbox"/>
8 Critère d'évaluation	<input type="checkbox"/>	<input type="checkbox"/>
9 Indicateur d'acquisition	<input type="checkbox"/>	<input type="checkbox"/>
10 Table de spécification	<input type="checkbox"/>	<input type="checkbox"/>
11 Évaluation formative	<input type="checkbox"/>	<input type="checkbox"/>
12 Évaluation sommative	<input type="checkbox"/>	<input type="checkbox"/>
13 Évaluation formatrice	<input type="checkbox"/>	<input type="checkbox"/>
14 Évaluation de l'état initial	<input type="checkbox"/>	<input type="checkbox"/>
15 Co-évaluation	<input type="checkbox"/>	<input type="checkbox"/>
16 Auto-estimation	<input type="checkbox"/>	<input type="checkbox"/>
17 Auto-évaluation	<input type="checkbox"/>	<input type="checkbox"/>
18 Pondération	<input type="checkbox"/>	<input type="checkbox"/>
19 Classement	<input type="checkbox"/>	<input type="checkbox"/>
20 Interlangue	<input type="checkbox"/>	<input type="checkbox"/>
21 Test	<input type="checkbox"/>	<input type="checkbox"/>
22 Taxonomie de Bloom	<input type="checkbox"/>	<input type="checkbox"/>
23 Validité	<input type="checkbox"/>	<input type="checkbox"/>
24 Fidélité	<input type="checkbox"/>	<input type="checkbox"/>
25 Théorie de l'apprentissage	<input type="checkbox"/>	<input type="checkbox"/>
26 Pédagogie par objectif	<input type="checkbox"/>	<input type="checkbox"/>
27 Performance observable	<input type="checkbox"/>	<input type="checkbox"/>
28 Objectif opérationnel	<input type="checkbox"/>	<input type="checkbox"/>
29 Outil d'évaluation fermé	<input type="checkbox"/>	<input type="checkbox"/>
30 Volume horaire d'apprentissage	<input type="checkbox"/>	<input type="checkbox"/>
31 Docimologie	<input type="checkbox"/>	<input type="checkbox"/>
TOTAUX	<input type="checkbox"/>	<input type="checkbox"/>

1. Vous trouverez les définitions de ces notions dans le lexique.

La terminologie liée au domaine de l'évaluation a pu être qualifiée de « jungle » par W. J. Popham².

Des essais de codification de ce vocabulaire ont été effectués par des chercheurs en éducation, mais de grandes disparités subsistent, selon les écoles, sur ce que recouvrent certains mots et concepts. Cela est dû d'une part aux difficultés liées à la traduction de ce lexique qui nous vient essentiellement de la recherche anglo-saxonne, et d'autre part de la qualité bien française de vouloir imposer des distinctions.

En voici deux exemples

- La traduction française du mot *skill*, utilisé par les Anglo-Saxons pour les quatre composantes de la communication langagière : expression et compréhension orales et écrites. Selon les ouvrages de référence, *skill* est traduit par « habileté » ou « capacité », parfois « compétence », voire « savoir-faire ». Or, comme on va le voir, ces quatre traductions recouvrent des concepts bien différents.

- Les deux mots « compétence » et « capacité » sont souvent employés l'un pour l'autre. Or, le concept de **capacité**, qui vient de l'anglais par *ability* ou *skill*, ne recouvre que le fait qu'une personne est capable de produire un comportement langagier donné dans une situation donnée, alors que le concept de **compétence** indique l'aptitude reconnue, **donc évaluée**, à savoir produire une conduite langagière

L'un des premiers objectifs de ce livre est de clarifier cette jungle terminologique, afin que les mêmes mots recouvrent toujours les mêmes concepts.

On trouvera à la fin un lexique sélectif commenté des mots et des concepts dont il sera question tout au long de l'ouvrage.

RÉFLEXION PRÉLIMINAIRE

Rares sont les pays où l'enseignant reçoit, dans sa formation initiale, ne serait-ce qu'une information sur les différentes façons d'évaluer les progrès en langues de ses élèves³.

Démuni, il reproduit souvent les procédures par lesquelles il a lui-même été évalué au cours de ses études. Cette évaluation scolaire, essentiellement dictée par les impératifs de l'institution scolaire ou universitaire, ne prend en compte

2. « Terminology jungle », cité par Louis Vandeveldé dans *Aider à devenir*, Bruxelles, Labor/Paris, Nathan, 1982, p. 149.

3. Voir à ce sujet, du même auteur, *La Classe de langue*, Paris, CLE International, coll. « Techniques et pratiques de classe », 1994.

Chapitre 1 • L'évaluation décortiquée

que l'un des aspects de l'évaluation, celle de l'évaluation sommative, qui permet de délivrer notes et diplômes, à usage social (passage d'une classe à l'autre, obtention d'un diplôme final de fin d'études).

Réfléchir à d'autres façons d'évaluer, c'est tout d'abord montrer du respect pour chaque individu, à partir de son plus jeune âge, pour ce qu'il est capable d'acquérir, pour les progrès qu'il fait dans son apprentissage. C'est également le moyen de lever un grand nombre de malentendus, entre élèves – qui comprennent difficilement les raisons des notes qu'ils obtiennent –, entre enseignants – qui s'évaluent implicitement entre eux –, entre enseignants et institutions, entre enseignants et parents, entre l'institution et la société dans son ensemble, en particulier avec le monde du travail.

De nombreux pédagogues réfléchissent à l'évaluation des acquis, en proposant des modalités d'évaluation différentes de la simple notation. Encore faut-il que la finalité de ces modalités ne soit pas d'aboutir *in fine* à une notation, car, en ce cas, l'évaluation censée être différente ne sera pas ressentie comme telle par les apprenants.

LES DEUX GRANDS TYPES D'ÉVALUATION

En évaluation des acquis, il faut tout d'abord séparer clairement tout ce qui relève du domaine du contrôle (évaluation sommative), de ce qui relève de l'information réciproque entre l'enseignant et l'apprenant (évaluation formative et auto-évaluation).

Le tableau de la page suivante permet de comparer ces deux grands types d'évaluation.

Le domaine du contrôle

► L'origine du mot « contrôle »

Le *Robert*⁵ indique qu'à l'origine le mot « contrôle » est formé de « contre » et de « rôle », le « contre rôle » désignant un registre de comptabilité (rôle) tenu en double, l'un servant à vérifier l'autre (contre). C'est cette fonction de vérification qui prévaut aujourd'hui, agrémentée de la notion de critique. Contrôler, c'est donc vérifier, en soumettant à critique, qu'un élément second correspond bien à un élément premier. En évaluation, contrôler signifie vérifier que la production d'un apprenant correspond bien à ce qu'édicte les normes grammaticales, lexicales, linguistiques.

5. *Dictionnaire historique de la langue française*, sous la direction d'Alain Rey, Paris, Dictionnaires Le Robert, 1998.

Le contrôle (évaluation sommative, normative)	La prise d'information (évaluation formative, critériée et auto-évaluation)
↓	↓
Contrôler, c'est vérifier la conformité des performances de l'apprenant à la	L'information que l'on recherche, c'est, autant pour l'apprenant que pour l'enseignant, de savoir si les objectifs fixés sont ou non atteints, selon les
↓	↓
norme de la langue cible.	critères formulés par l'enseignant.
↓	↓
Pour cela, on utilise des tests calibrés, des exercices, des examens, qui donnent lieu à une	Pour cela, on utilise des activités d'évaluation, des fiches d'auto-estimation, des exercices de vérification qui donnent lieu à une
↓	↓
note Cette évaluation est imposée , elle sert à classer les élèves entre eux. La note est la seule information donnée. Il s'agit d'une évaluation sanction.	information commentée Cette évaluation est consentie , elle apporte des informations sur les acquis et sur ce qu'il reste à acquérir. Il s'agit d'une évaluation (in)formative.
↓	↓
Elle mène à la certification. Elle a une valeur sociale.	Elle mène à la reconnaissance des compétences. Elle a une valeur formative.

Les moyens mis à la disposition de l'enseignant pour contrôler les acquis sont multiples: ils vont des activités de classe, notées ou non, aux examens terminaux.

► La norme

Elle est opposée aux critères. Ni l'enseignant, ni l'apprenant n'ont de prise sur elle. La norme est définie *a priori*: c'est celle des grammaires, des dictionnaires, des manuels d'apprentissage, du système phonologique.

Il est normal que l'enseignant tente de faire acquérir cette norme, car son souhait est que ses élèves s'en rapprochent le plus possible pour « parler correctement ». Un individu qui respecterait la totalité de la norme pourrait être qualifié de bilingue. Or la finalité de l'enseignement n'est pas de fabriquer des bilingues mais bien de conduire les apprenants du niveau A où on les prend en charge au niveau B qui sera l'objectif de la période d'apprentissage. La recherche du bilinguisme est un fantasme qui n'est pas de mise dans les classes de langue.

Chapitre 1 • L'évaluation décortiquée

Le *Cadre européen commun*, en définissant des niveaux de compétence en langue, indique en quelque sorte des paliers d'apprentissage. L'enseignant et les systèmes éducatifs doivent admettre que l'objectif du système n'est pas de faire acquérir aux élèves, au cours de leur scolarité, le palier le plus haut (le niveau C2 du *Cadre*). Des paliers progressifs doivent être établis pour chaque étape de l'apprentissage, et c'est uniquement la maîtrise de ces paliers – de ces niveaux – qui doit être évaluée. Il faut donc aujourd'hui savoir situer la norme. Des outils conceptuels tels que le *Cadre européen commun* et les *Référentiels* qui s'y rattachent sont pour cela inestimables. Tant que ce concept ne sera pas assimilé, la norme inatteignable continuera à être la référence des enseignants et la hantise des apprenants.

► **La note en situation de classe, la fidélité de la notation, le mythe entretenu par la docimologie**

D'où vient l'idée d'attribuer des notes aux travaux et aux performances des élèves ? Qu'implique cette note ? Peut-on considérer que chaque performance puisse se voir attribuer une note réellement représentative de la valeur intrinsèque du travail fourni ? Pourquoi les notes diffèrent-elles d'un correcteur à l'autre ? Pourquoi un même travail peut-il être évalué différemment, à quelques mois ou semaines d'intervalle, par un même correcteur ?

Les travaux de Piéron⁶ et d'autres chercheurs, nés du constat des différences individuelles intra- et inter-évaluateurs, ont proposé dans les années 1960 une nouvelle discipline, la docimologie (étude critique des examens et des concours), pour montrer qu'il était possible d'obtenir pour chaque travail fourni une « vraie note » représentative de sa valeur. La docimologie considérait que les travaux (copies d'élèves par exemple, lors des examens sommatifs) pouvaient être mesurables et quantifiables. Les différences de notation pouvaient être éliminées en multipliant le nombre de correcteurs et en appliquant aux diverses notes obtenues une procédure statistique de mesure permettant de déterminer la « vraie note ».

Des chercheurs⁷ ont calculé, dans différentes disciplines enseignées à l'école primaire puis au niveau du baccalauréat (français, mathématiques, philosophie, physique), le nombre de correcteurs au-delà duquel la moyenne des notes obtenue ne variait plus : 128 correcteurs en philosophie, 78 en français, 16 en physique, 13 en mathématiques. Ces résultats étaient logiques, mais coûteux, inapplicables et en quelque sorte malheureux : ils ont donné naissance au mythe de la « vraie note », que tout enseignant aimerait pouvoir attribuer aux performances de ses élèves de façon à constater le plus ou moins grand écart avec la norme.

6. H. Piéron, *cf.* Bibliographie.

7. H. Laugier et D. Weinberg, *cf.* Bibliographie.

Il n'a fallu qu'une dizaine d'années pour que les résultats des travaux issus de la docimologie soient remis en question⁸.

Une performance orale ou écrite (sauf si l'épreuve se présente sous forme de questions à choix multiples – QCM – standardisées) ne peut être mesurée comme un objet physique concret, à l'aune d'un instrument de mesure. La notation, sauf dans le cas de tests calibrés⁹, n'est pas un instrument de mesure exacte, elle ne contribue pas à la fidélité de l'évaluation. Elle ne peut être qu'une estimation donnée à une performance, en fonction de critères d'appréciation.

La docimologie a cependant le mérite d'avoir montré que la notation de copies n'est jamais stable ni constante (les qualités essentielles de la fidélité) et que de nombreux facteurs, souvent connus sous le terme de « critères parasites de la fidélité », y contribuent.

Tout enseignant connaît les critères parasites de la fidélité, sans toujours savoir les nommer. Ils sont nombreux, il faut savoir les reconnaître pour en prendre conscience et essayer de les réduire, voire de les éliminer.

- *L'effet de fatigue* : les conditions de correction influencent le jugement de l'évaluateur. On corrige sans aucun doute plus équitablement le matin, bien assis à une table de travail que le soir dans les transports en commun...

- *L'effet de contraste* : une copie moyenne souffre d'être corrigée juste après une excellente production. Si elle avait été corrigée juste après une copie très médiocre, il y a fort à parier qu'elle aurait obtenu une bien meilleure note...

- *L'effet de favoritisme* : le favori ou le « chouchou » est en général celui qui renvoie à l'enseignant une image gratifiante de lui-même. Quel dommage de lui mettre une « mauvaise » note...

- *L'effet d'ordre* : on est en général malheureusement plus sévère à la fin d'une série de copies qu'au début. La note attribuée à un travail dépend donc de sa place dans le paquet.

- *L'effet de stéréotypie* : les notes attribuées en début de formation constituent parfois une référence pour l'enseignant qui, par la suite, a du mal à noter différemment. Combien d'élèves disent qu'ils ont été jugés une fois pour toutes et que leurs notes ne varient pas quels que soient les efforts qu'ils fournissent ?

- *L'effet de contamination* : un bon dossier scolaire ne rachète pas une copie faible; en revanche il influe favorablement sur une bonne copie. Cet effet a été démontré « à l'aveugle » : une copie moyenne a été présentée à l'évaluateur comme venant d'un élève faible et à d'autres comme venant d'un bon élève. Dans le premier cas, elle a été sous-évaluée, et surévaluée dans le second cas.

8. J.-J. Bonniol, cf. Bibliographie.

9. Cf. « Calibrage » dans le lexique.