

NE PAS SE TROMPER D'ÉVALUATION

Jean-Marie De Ketele

Pub. linguistiques | « [Revue française de linguistique appliquée](#) »

2010/1 Vol. XV | pages 25 à 37

ISSN 1386-1204

Article disponible en ligne à l'adresse :

<https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2010-1-page-25.htm>

Distribution électronique Cairn.info pour Pub. linguistiques.

© Pub. linguistiques. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

Ne pas se tromper d'évaluation

Jean-Marie De Ketele

Université Catholique de Louvain, Louvain-la-Neuve (Belgique)

Résumé : *Cet article part du constat curieux que, dans un monde où les métiers du complexe se professionnalisent, l'évaluation scolaire se résume encore trop souvent à l'attribution d'une note. Conduite professionnellement, l'évaluation suppose que l'évaluateur se pose avant tout les questions relatives aux types de fonctions à viser et aux types de démarches à mettre en œuvre. Plus spécifiquement, l'article aborde deux aspects importants de la professionnalisation : l'évaluation des compétences langagières et le renforcement des fonctions d'orientation et de régulation.*

Abstract: *This article starts from the curious observation that, in a world where complex trades become increasingly professionalized, assessments in schools come generally down to awarding marks. If done professionally, the evaluator should before all other things answer questions about the types of functions that the evaluation has to fulfill and the procedures to follow. More specifically, the article addresses two important aspects of professionalization : evaluation of language competencies and the reinforcement of the functions of orientation and regulation.*

Pour l'homme de la rue, évaluer est un processus qui consiste à attribuer une note. C'est là une conception très restrictive qui montre que les travaux des experts en évaluation ont bien du mal à contrecarrer des représentations solidement ancrées dans nos histoires de vie : les notes scolaires ont été pour la plupart d'entre nous des déclencheurs de sentiments divers – des joies, des tristesses, le sens du devoir accompli ou des efforts récompensés, des sentiments d'injustice passagers ou parfois ravageurs à long terme, etc. Il est curieux qu'à une époque où les métiers de l'humain manifestent des efforts considérables pour se professionnaliser, les enseignants véhiculent encore eux-mêmes une telle conception et des pratiques aussi uniformes. C'est d'autant plus curieux, qu'à l'instar des professions libérales, ils travaillent dans le complexe et que leurs activités réclament de fait un corps de connaissances, de techniques et d'expériences pratiques que seule une formation longue (initiale) et jamais inachevée (formation continue) permet de leur donner (Maroy 2006). Et on peut s'étonner d'autant plus que, dès les années 50, les docimologues avaient tenté d'attirer l'attention des enseignants sur les biais liés au processus d'attribution d'une note (Piéron 1963 ; de Landsheere 1974 ; De Ketele 1985) et que, par la suite, les chercheurs en psychologie sociale montraient combien les processus évaluatifs devaient être conduits en connaissance de cause, tant les effets pouvaient se révéler puissants et redoutables (Gilly 1980 ; Noizet et Caverni 1978). Dans un esprit plus littéraire, Pennac (2007), fort de son ancien statut de « cancre », montre bien ses effets pervers dans *Chagrin d'école*.

On peut se demander si les résultats de chercheurs en docimologie, malgré leur intérêt indéniable, n'ont pas contribué à ancrer encore plus dans les esprits une conception très étroite du processus évaluatif. Puisque, étymologiquement, la docimologie est la science des examens et que les travaux des chercheurs ont porté essentiellement sur les notes comme

échelles de mesure des performances, trop de praticiens en ont déduit que l'évaluation se résumait à attribuer une note de façon fiable pour sanctionner la réussite des apprentissages. C'est là réduire l'évaluation à une seule fonction (« sanctionner »), une seule démarche (« faire une somme ») et une seule finalité (« porter un jugement »).

Or, depuis plus de cinquante ans, les travaux des chercheurs en évaluation ont bien montré qu'il existe plusieurs fonctions possibles de l'évaluation, que pour chacune des fonctions plusieurs démarches sont envisageables avec des effets différents et que la finalité de l'évaluation, si on se réfère aux travaux initiés par Stufflebeam & al. (1980) (voir la revue de la littérature de De Ketele (1993) dans la *Revue Française de Pédagogie*), est de fonder des prises de décision et non de porter des jugements sur les personnes. Illustrons ces aspects dans le domaine de l'évaluation des apprentissages langagiers.

1. Les fonctions de l'évaluation

On peut distinguer au moins trois fonctions principales de l'évaluation : évaluer pour orienter une nouvelle action à entreprendre ; évaluer pour améliorer une action en cours ; évaluer pour certifier socialement une action considérée comme terminée.

C'est cette troisième fonction à laquelle se réfèrent, malheureusement pas toujours de façon pertinente, les enseignants. Rigoureusement parlant, l'**évaluation certificative** consiste à certifier socialement (devant les autorités institutionnelles, les parents, les enfants et les collègues) les effets d'une action menée et considérée comme achevée. Si on se réfère au contexte de l'école, elle vise à prendre la décision suivante : « Au terme de..., nous certifions que l'élève X a réussi ou échoué ou atteint tel degré de performance et qu'il a donc les acquis suffisants en ... pour... ». Les points de suspension sont à préciser en fonction du contexte. Ainsi, en fin de telle année scolaire, la décision pourra être formulée comme suit (nous verrons ultérieurement que ce n'est pas la seule formulation possible mais que la structure de la formulation est cependant la même) : « Au terme de cette année scolaire (ex. : CM1 ou CM2), nous certifions que l'élève X a réussi son année scolaire et maîtrise les acquis langagiers nécessaires et suffisants pour commencer avec fruit l'année scolaire suivante (CM2 ou 6^e) ». Plusieurs commentaires importants méritent d'être faits : (1) la certification est toujours un processus social qui engage les évaluateurs qui ont pris la décision (c'est tellement vrai que nous assistons de plus en plus à des recours, y compris en justice ; on parle de « judiciarisation » de l'enseignement) ; (2) la certification de fin d'année ou de fin de cycle suppose que les évaluateurs aient identifié suffisamment clairement les acquis langagiers nécessaires et suffisants pour passer d'une année dans une autre année (le problème se complique quand au terme d'une année plusieurs filières différentes sont offertes). Les certifications en fin d'année ou fin de cycle sont évidemment les plus importantes ; ceci ne veut pas dire qu'il ne peut exister de certifications intermédiaires, comme par exemple des évaluations de fin de trimestre, voire de périodes plus courtes. Mais la fonction certificative exige que la séquence d'apprentissage portant spécifiquement sur l'aspect langagier évalué soit terminée et non en cours et que l'évaluateur n'oublie pas que l'élève peut continuer à apprendre (ou à désapprendre) entre l'évaluation intermédiaire et l'évaluation finale. Dans leurs pratiques évaluatives, beaucoup d'enseignants oublient cette double exigence et commettent ainsi des « échecs abusifs » : en intégrant dans leurs évaluations certificatives des « interrogations » ou des « contrôles » sur des apprentissages en cours et donc non terminés, ils pénalisent les élèves qui ont encore besoin de l'enseignant et sélectionnent indirectement les élèves qui peuvent s'en sortir sans le professionnalisme de l'enseignant. Notons que l'évaluation certificative dans le domaine du langage est à ce niveau bien plus complexe que dans le domaine des disciplines plus linéaires : il est bien plus facile de considérer qu'une

séquence d'apprentissage sur la solution d'une équation du second degré est terminée, que pour une séquence d'apprentissage sur le résumé d'un texte argumentatif.

Comme nous venons de le laisser entendre, une des difficultés principales rencontrées dans les pratiques est la confusion entre la fonction certificative et **la fonction de régulation** (mieux connue dans le monde de l'école sous l'appellation d'évaluation formative). Rigoureusement parlant, l'évaluation de régulation (ou formative) est un processus qui doit permettre de prendre des décisions pour améliorer l'action (c'est-à-dire ici l'apprentissage) en cours (la séquence d'apprentissage concernée n'est pas terminée). Les objets de l'évaluation de régulation seront le plus souvent des objets plus circonscrits que les objets de l'évaluation certificative. Cette dernière sera par exemple centrée sur la production de messages alors que l'évaluation de régulation portera sur les « ressources » à acquérir pour les produire, comme des règles linguistiques ou comme l'identification des caractéristiques essentielles du type de message à produire. L'évaluation de régulation est donc centrée sur l'identification de ce qui est déjà acquis ou non encore acquis à un moment donné d'un apprentissage non terminé afin de prendre des décisions pour améliorer l'apprentissage en cours. L'identification des erreurs, l'émission d'hypothèses explicatives sur leurs sources, l'anticipation d'actions pour y remédier sont alors au cœur du processus évaluatif : c'est sans doute dans l'exercice de la fonction de régulation que le professionnalisme de l'enseignant se révèle le plus. Alors que dans l'évaluation certificative, l'erreur est sanctionnée (du moins si elle est fondamentale) puisque l'apprentissage est censé être terminé (nous sommes « au terme de... »), dans l'évaluation de régulation le statut de l'erreur change puisque son identification et le travail sur celle-ci est moteur de l'apprentissage.

Si le concept d'évaluation formative est connu de tous les enseignants (mais dans les pratiques souvent confondu avec l'évaluation certificative), le concept **d'évaluation d'orientation** l'est beaucoup moins, mais est néanmoins très important. Rigoureusement parlant, l'évaluation d'orientation a pour fonction de préparer une nouvelle action : une nouvelle année ou une nouvelle filière d'études pour un élève ; une nouvelle année scolaire pour un enseignant ou un établissement ; un nouvel enseignement ; une nouvelle séquence d'apprentissage... Dans toutes ces situations, les enjeux sont importants et méritent que les décisions qui seront prises soient fondées sur un processus d'évaluation pertinent. Ainsi, l'orientation d'un élève au terme du collège a fait couler beaucoup d'encre en France et certains¹ doutent que la seule prise en considération des notes scolaires antérieures soit entièrement pertinente. De même, De Ketele (1996) a montré que pas plus de 30 % de la performance académique des étudiants de première année d'université était explicable par les facteurs d'aptitude cognitivo-scolaire. Ainsi encore, une série d'études répétées à travers le temps (De Ketele 2006) ont montré que les résultats à un test national (tant en français qu'en mathématiques) passé en fin de scolarité primaire (fin juin) et repassé aux mêmes élèves au début du collège (début septembre) montraient une déperdition moyenne de 25 % ; suite à ces résultats, des études ont montré que les enseignants, mettant en place au début de l'année scolaire une évaluation portant sur les prérequis spécifiques des nouveaux apprentissages prévus (il s'agit bien d'une évaluation d'orientation) et organisant des moments de consolidation de ceux-ci, obtenaient ensuite de meilleurs résultats que ceux commençant directement les nouveaux apprentissages (De Ketele 2006). On pourrait encore faire état des enseignants ou des équipes d'enseignants, avec ou sans l'aide des responsables

¹ Voir, par exemple, l'article du *Nouvel Observateur* du jeudi 8 avril 2010 où Bruno Suchaut s'étonnait de voir la façon dont l'orientation était assurée sur la base des notes scolaires. Dans une communication au Colloque de Luxembourg (2006) de l'Association pour le Développement des Méthodologies de l'Évaluation en Éducation, le même auteur montrait sur la base d'une évolution temporelle des résultats aux évaluations nationales de français au CE2 que les variations temporelles de notes ne signifiaient pas des baisses ou des hausses de niveau des élèves. La démonstration pourrait être faite pour d'autres niveaux d'études.

d'établissement, qui préparent une nouvelle année scolaire sur la base d'une analyse d'un ensemble de données relatives aux expériences diverses vécues l'année précédente ou pour lesquelles des traces ont été gardées à de telles fins. Nous sommes bien là dans des processus d'évaluation d'orientation qui permettent de fonder des prises de décision pour de nouvelles actions.

2. Les démarches

Il ne faut pas confondre les fonctions et les démarches de l'évaluation. La fonction se réfère au « pour quoi » de l'évaluation (préparer, améliorer ou certifier), tandis que la démarche se réfère au « comment ». L'analyse des pratiques montre que la seule démarche connue, ou du moins utilisée, par la plupart des évaluateurs du terrain scolaire est la démarche sommative (mettre une note, c'est-à-dire faire une somme), à tel point que, pour la plupart d'entre eux, l'évaluation sommative est assimilée à l'évaluation certificative. A la décharge des enseignants, il faut reconnaître que les experts en évaluation des années 1960 (au départ des travaux de Scriven 1967) ont répandu la distinction entre l'évaluation sommative et l'évaluation formative. S'il s'agissait d'un pas en avant dans le domaine de l'évaluation, cela a néanmoins constitué une erreur historique qui a été répandue largement dans les formations initiale et continue des enseignants : le concept d'évaluation formative relève bien de la fonction (pour améliorer l'enseignement en cours) tandis que le concept d'évaluation sommative relève bien de la démarche (le comment, c'est-à-dire ici faire une somme pour produire le résultat de l'évaluation). Dans les faits, il existe trois grands types de démarches possibles pour évaluer et celles-ci peuvent être théoriquement utilisées pour chacune des trois fonctions.

2.1. La démarche sommative consiste à sommer (au sens algébrique du terme), c'est-à-dire additionner des éléments (j'additionne des points censés représenter des niveaux de performance) ou soustraire des éléments (je soustrais des points d'un maximum en fonction des erreurs), en tenant compte, consciemment ou inconsciemment, des pondérations accordées aux éléments. Contrairement à beaucoup d'autres domaines où l'on somme des connaissances nouvelles apprises, celui des langues utilise principalement une démarche sommative soustractive, du moins en ce qui concerne la maîtrise des ressources linguistiques : l'évaluateur part d'une production correcte sur le plan linguistique et opère une soustraction chaque fois qu'il rencontre une incorrection. Lorsqu'il s'agit d'une production complexe (une composition française, une dissertation ou une analyse littéraire), l'évaluateur utilise le plus souvent une double démarche, soustractive pour la maîtrise linguistique et additive pour le fond. Ceci n'est pas sans soulever des problèmes de mesure dont nous aurons l'occasion de reparler, surtout si la fonction de l'évaluation est certificative. Si la fonction est formative (réguler l'apprentissage en cours) ou d'orientation (préparer une nouvelle action), on peut se demander si la démarche sommative est la plus indiquée dans le contexte précis où elle est appliquée.

2.2. La démarche descriptive consiste à identifier et décrire les éléments ou les caractéristiques de l'objet à évaluer. Si la fonction est certificative, il s'agira non pas de donner une note ou un profil de notes, mais d'identifier et de décrire les acquis de l'élève à un moment donné de son parcours : l'évaluateur ou les évaluateurs certifient devant les instances concernées que l'élève a acquis, à un seuil de maîtrise suffisant et de façon suffisamment stable, les ressources et les compétences décrites dans le certificat. Dans certains mouvements pédagogiques ou dans certaines communautés scolaires, on n'utilise pas (et on ne veut pas utiliser) des notes, mais on s'attache à décrire progressivement les acquis des élèves et on parlera de « bulletins descriptifs ». Cette pratique est particulièrement rejetée par beaucoup

d'enseignants (le pouvoir de la note ; la note comme instrument de motivation extrinsèque) et beaucoup de parents (le sentiment d'être dépassé par la signification des formulations employées ; le besoin de savoir le classement de leurs enfants). Dans les formations plus professionnalisantes, on utilise de plus en plus des certificats descriptifs, sous l'impulsion des employeurs qui n'ont pas confiance dans les notes attribuées, mais ont besoin avant d'engager du personnel de connaître les compétences qui ont été certifiées. Si la fonction est de régulation ou formative, la démarche descriptive est plus intéressante puisque, pour améliorer l'apprentissage en cours, il s'agit de pouvoir identifier et décrire les difficultés des élèves et de leur donner du sens par rapport aux apprentissages antérieurs et par rapport aux objectifs visés. Il en va de même dans bien des cas pour la fonction d'orientation, car une note ou un ensemble de notes peut rarement prédire le succès d'une orientation et il faut pouvoir identifier et décrire les facteurs relevant d'une orientation réussie.

2.3. Moins connue, quoique pratiquée dans bien des lieux et des circonstances, est le troisième type de démarche. Son appellation n'est d'ailleurs pas encore bien stabilisée. Nous aimons l'appeler « **démarche herméneutique** » (étymologiquement, chercher le sens caché derrière les relations entre les phénomènes observés), mais elle est également appelée selon les auteurs « démarche interprétative » ou « démarche heuristique » (Jorro 2007) ou « démarche clinique » (cf. les travaux de Piaget) ou même « démarche intuitive » (voir Weiss 1984). Quel que soit le nom donné, la démarche obéit à quatre caractéristiques essentielles : (i) le recueil d'un ensemble d'indices de nature diverse (indices quantitatifs ou/et qualitatifs, de sources variées) ; (ii) une tentative pour organiser les indices en un ensemble cohérent qui donne sens ; (iii) un processus de validation d'une hypothèse parmi plusieurs hypothèses alternatives ; (iv) la démarche globale sous forme d'un raisonnement (de nombreux professionnels parlent de « raisonnement clinique »). Cette démarche est plus fréquente qu'il n'y paraît. Ainsi, la certification de la réussite de nombreux stages (notamment ceux des futurs enseignants) se fait essentiellement par une démarche herméneutique (même si l'institution demande d'engranger une note finale !). De nombreuses décisions pour améliorer un apprentissage en cours se font également par un processus herméneutique. Ceux qui ont à préparer une nouvelle action importante rassemblent de nombreux indices jugés pertinents (quantitatifs et qualitatifs) auxquels ils donnent sens ; la qualité de l'orientation choisie dépendra de la qualité du raisonnement qui sera opéré sur de tels indices.

3. Le croisement des fonctions et des démarches

Le croisement des trois fonctions et des trois démarches fait apparaître un éventail de 9 processus différents d'évaluation. Pour chacun de ces processus, nous donnons un exemple appliqué au domaine de l'apprentissage dans le domaine des langues (cf. tableau 1).

Ce croisement nous permet de mettre en évidence trois confusions qui prennent leur source dans l'évolution historique des travaux sur l'évaluation. La première, que nous avons déjà évoquée, tient aux travaux de Scriven (1967) qui a montré qu'il existait au moins deux types de processus évaluatifs : (i) attribuer une note pour « sanctionner » le résultat d'un apprentissage ; (ii) prendre et analyser de l'information pour améliorer l'apprentissage en cours. Il s'agissait bien d'une première avancée importante dans le domaine de l'évaluation, mais qui a induit la confusion entre fonction et démarche ; l'évaluation certificative était assimilée à l'évaluation sommative, alors que nous venons de montrer que la fonction certificative pouvait être conduite avec trois types de démarches.

Tableau 1. *Le croisement fonctions x démarches : quelques exemples.*

Fonctions→ Démarches ↓	Orientation <i>Préparer une nouvelle action</i>	Régulation <i>Améliorer une action en cours</i>	Certification <i>Certifier le résultat d'une action</i>
Sommative <i>Mettre une note</i>	Ex. 1 : Attribuer une note à un test de prérequis langagiers Ex. 2 : Orienter un étudiant sur la base d'un dossier de notes scolaires (comme au terme du collège)	Ex. 1 : Attribuer une note à une interrogation en cours d'apprentissage sur les règles d'accord Ex. 2 : Noter sur 10 points un « quizz test » de 10 questions de compréhension sur un texte en cours d'analyse	Ex. 1 : Attribuer une note à une composition française Ex. 2 : Noter en fin d'année un test composite (analyse, conjugaison...)
Descriptive <i>Identifier et décrire</i>	Ex. 1 : Identifier parmi 10 prérequis de la lecture, ceux qui sont déjà maîtrisés par un élève du préscolaire Ex. 2 : Identifier dans les questions d'examen de français de l'an dernier les questions qui posent problème	Ex. 1 : Repérer dans la classe les élèves qui n'ont pas réussi un exercice individuel de conjugaison pour organiser un tutorat Ex. 2 : Organiser un test diagnostique sur un apprentissage important (voir technique présentée dans le texte)	Ex. 1 : Remplir un bulletin descriptif (cocher dans une liste d'objectifs du cours de français ceux qui sont atteints de façon stable) Ex. 2 : Rédiger un certificat mentionnant et décrivant les compétences langagières acquises
Herméneutique <i>Donner sens</i>	Ex. 1 : Orienter un étudiant vers une filière littéraire sur la base de nombreux indices (résultats scolaires, observations, entretien) Ex. 2 : Analyser finement en début d'année une première dictée et un premier texte libre des nouveaux élèves en vue de préparer les premières séquences	Ex. 1 : Observer et analyser les productions des élèves pendant une séquence de travail individuel en classe Ex. 2 : Analyser les résultats d'un test diagnostique à la lumière d'une analyse de la façon dont les séquences d'apprentissage antérieures ont été réalisées	Ex. 1 : Préparer une réunion de parents sur la base d'un portfolio des productions significatives de ses élèves Ex. 2 : Certifier le passage de classe, non seulement sur la base des notes obtenues en fin d'année, mais en tenant compte d'autres indices pertinents

La seconde prend son origine dans une seconde évolution historique des travaux en évaluation : Bloom & al. (1971) a complété la distinction de Scriven (évaluation sommative + évaluation formative) par un troisième type d'évaluation, l'évaluation diagnostique désignant par là l'évaluation qui vise à préparer une nouvelle action. C'était un pas en avant important, mais l'appellation était malheureuse, car il y a autant de processus diagnostiques dans l'évaluation formative ou de régulation que dans l'évaluation d'orientation ; ainsi, un test diagnostique peut aussi être réalisé pour l'une ou l'autre de ces deux fonctions. C'est pourquoi le concept d'évaluation d'orientation est plus approprié pour désigner cette nouvelle avancée. Une troisième confusion, encore très ancrée et source de nombreux « dégâts », tient à la finalité du processus évaluatif. Si tous les auteurs actuels s'accordent sur le fait que le processus évaluatif consiste à confronter un référent à un référé, ils divergent sur la finalité du processus : pour les uns, il s'agit de « porter un jugement sur » ; pour les autres, de « fonder une prise de décision ». Sur la base des travaux de Stufflebeam & al. (1980) et d'autres auteurs (De Ketele & Roegiers 1991), nous pensons que le processus évaluatif sert à fonder une prise de décision, soit pour préparer une nouvelle action (fonction d'orientation), soit pour améliorer une action en cours (évaluation formative ou de régulation), soit pour certifier

socialement le résultat d'une action (évaluation certificative). Ceci ne veut pas dire que le processus évaluatif n'est pas exempt de processus de jugement, mais ceux-ci se trouvent à certaines étapes du processus évaluatif et non comme finalité (l'évaluation ne peut avoir pour finalité de juger une personne). C'est ce que nous traduisons dans la définition suivante que nous voulons la plus opérationnelle possible et que nous présentons sous forme décomposée afin de mettre en évidence où des processus de jugement et d'attribution de valeurs sont présents :

<p>Le processus évaluatif =</p> <ul style="list-style-type: none"> - recueillir un ensemble d'informations (<i>choix = processus de jugement</i>) <ul style="list-style-type: none"> o pertinentes dans le choix, o valides pour l'exploitation o fiables dans le recueil - et confronter, par une démarche adéquate, cet ensemble d'informations à un ensemble de critères (<i>révélateur du système de valeurs</i>) <ul style="list-style-type: none"> o pertinents dans le choix (cohérent avec le référentiel) o valides dans l'opérationnalisation o fiables dans l'utilisation - pour attribuer une signification aux résultats de cette confrontation (<i>processus d'attribution de sens</i>) - et ainsi pouvoir fonder une prise de décision cohérente avec la fonction visée par l'évaluation (<i>orientation vers l'action à prévoir, en cours ou achevée</i>)
--

Une telle présentation nous permet d'insister sur plusieurs points qui nous semblent particulièrement importants. Si, avec raison, la docimologie a mis en évidence l'importance de la fiabilité lorsqu'on adopte une démarche sommative (attribution d'une note), il ne faut pas oublier que, quelle que soit la démarche adoptée, d'autres qualités sont nécessaires, tant dans le recueil que dans le traitement de l'information. Etre pertinent, c'est répondre à la question « est-ce que je ne me trompe pas de... ? » : ne pas se tromper d'informations à recueillir, ne pas se tromper de méthode de recueil (un QCM et un test de production ne sont pas équivalents ; dois-je utiliser un oral ou un écrit ?...), ne pas se tromper de critères (avons-nous explicité et discuté les critères d'une dissertation pour des étudiants de tel niveau ?)... Etre valide, c'est répondre à la question « ce que je déclare faire et ce que je fais réellement correspondent-ils ? » : ainsi, à titre de contre-exemple, je déclare évaluer la qualité du résumé produit par l'élève et j'évalue en fait la qualité du résumé fait par l'enseignant et restitué par l'élève... Etre fiable, c'est répondre à la question « ce que j'obtiens comme information recueillie ou comme résultat du traitement est-il bien ce qui aurait pu être obtenu par un autre évaluateur ou dans une autre circonstance ? » : au-delà des biais établis par les docimologues à propos de l'attribution d'une note, on n'oubliera pas non plus les biais dans la sélection de l'information (qu'est-ce que je regarde dans ce que l'élève me montre ?) et dans l'utilisation des critères (les critères ont-ils le même sens pour deux évaluateurs différents ?)... Tout ceci montre bien qu'à chaque étape du processus d'évaluation, il y a des choix à effectuer qui dénotent le système de valeurs de l'évaluateur (évaluer, c'est attribuer de la valeur), mais qu'au terme du processus on ne peut se contenter d'un jugement sur l'évalué et qu'il s'agit de prendre les décisions adéquates en cohérence avec les jugements de valeur effectués tout au long du processus. Ceci nous permet aussi de souligner l'importance de donner du sens au résultat du traitement de l'information. Si l'attribution de sens est au cœur de la démarche herméneutique, elle ne l'est pas naturellement dans la démarche sommative : ainsi, deux notes de 10/20 données à une production langagière peuvent signifier des performances très différentes et de ce fait induire des décisions différentes pour l'action.

4. Un problème actuel majeur : l'évaluation des compétences

La distinction entre évaluation des ressources et évaluation des compétences est désormais largement répandue. Illustrons-la à travers quelques questions d'évaluation dans le domaine langagier :

<p>RESSOURCES</p> <ol style="list-style-type: none"> Démarche de restitution d'un connaissance <ul style="list-style-type: none"> Ex.1 : Énoncez la règle d'accord du verbe avec le sujet. Ex.2 : Citez plusieurs types de textes selon la fonction visée. Démarche d'application scolaire <ul style="list-style-type: none"> Ex.1 : Accordez les phrases suivantes (phrases dont le verbe est donné à l'infinifitif). Ex.2 : Lisez le texte suivant et dites de quel type de texte il s'agit. Démarche d'application habillée <ul style="list-style-type: none"> Ex.1 : Transformez l'invitation suivante qui s'adresse à votre ami Paul pour qu'elle s'adresse à la fois à votre ami Paul et à sa sœur Virginie (l'invitation à une fête est donnée et est adressée à Paul). Ex.2 : Le texte suivant est un texte argumentatif sur la pollution. Adaptez-le pour qu'il devienne un texte purement descriptif (texte argumentatif donné). <p>COMPÉTENCES</p> <ol style="list-style-type: none"> Production très contextualisée <ul style="list-style-type: none"> Ex.1 : Voici un texte de Victor Hugo racontant l'aventure vécue par son héros et dont il manque le 3^e paragraphe. En respectant l'histoire (les deux premiers paragraphes) et le style de Victor Hugo, rédigez le 3^e paragraphe. Ex.2 : Rédigez en une page maximum, pour la revue de l'école, la rencontre que nous avons vécue hier avec un ancien de l'école. Production très libre <ul style="list-style-type: none"> Ex.1 : Commentez : « Une idéologie est toujours politique ». Ex.2 : Rédigez pour la revue de l'école un article de votre choix.
--

Dans cet encadré, nous avons illustré différents niveaux de « savoir » et tenté de montrer le saut qualitatif entre la maîtrise des ressources et la maîtrise des compétences. Nous tenons à insister sur la confusion fréquente (non seulement chez beaucoup d'enseignants, mais même dans des épreuves nationales et internationales) entre l'application habillée et la compétence : dans l'application habillée, l'énoncé de la question dit sans ambiguïté les connaissances ou savoir-faire à mobiliser (mettre l'invitation au pluriel puisqu'il y a Paul et Virginie ; supprimer les indices argumentatifs du texte) ; l'énoncé visant à évaluer une compétence ne dit pas quels connaissances ou savoir-faire doivent être mobilisés (il faut analyser l'histoire et dégager le style de Victor Hugo pour effectuer la tâche complexe demandée ; les productions de textes, contextualisées ou davantage libres, supposent un travail sur la recherche des connaissances ou savoir-faire à mobiliser).

Au plus les questions d'évaluation se réfèrent à des tâches complexes, au plus l'évaluation se révèle complexe et demande du professionnalisme. La tentation est de se réfugier dans un technicisme étroit qui enlève le sens à donner à la production de l'élève ou qui amène à faire des choix non pertinents.

En effet, que signifie une note de 10/20 donnée à une compétence langagière (cf. les exemples de compétences dans l'encadré), note attribuée en sommant deux notes partielles, l'une sur la « forme » et l'autre sur le « fond » ? Notons positivement d'emblée qu'une telle

proposition souligne que la compétence est multidimensionnelle (ici deux dimensions) et que l'on pourrait très bien avoir une excellente maîtrise de la forme et non du fond ou inversement. Ainsi un 10/20 peut être le fruit de 22 combinaisons de performance différentes (de 10 pour la forme + 0 pour le fond ... à 0 pour la forme + 10 pour le fond). La forme et le fond sont deux dimensions théoriquement indépendantes et, même s'il est fréquent de constater que les élèves forts sur une dimension le sont également sur l'autre, la pratique montre que tous les cas de figures existent (Gerard 2008) et que l'évaluateur éprouve des difficultés à respecter la règle de l'indépendance des critères d'évaluation et se laisse influencer par le biais « effet halo ». Plus fondamentalement encore se pose le problème de mesure : rigoureusement parlant, on ne peut additionner sur une échelle unidimensionnelle les scores obtenus sur des échelles de mesure de dimensions différentes théoriquement indépendantes. La sagesse voudrait que l'on utilise, non pas une note unique, mais un vecteur de notes pour souligner le caractère multidimensionnel de la compétence, ce qui permet de donner une signification plus valide à la performance.

Le mécanisme de défense qui consiste à croire qu'une haute technicité résout tous les problèmes est fréquent chez certains experts en évaluation. Ainsi, pour être validées, de nombreuses épreuves d'évaluation nationales et internationales recourent à des modèles mathématiques très sophistiqués, dont le plus connu est le Modèle TRI (Théorie de Réponse à l'Item). De Ketele et Gerard (2005) montrent que ce modèle fonctionne de façon pertinente pour l'évaluation de tâches peu complexes (celles ayant trait aux ressources) mais non pour l'évaluation de tâches complexes (c'est-à-dire des compétences, et tout particulièrement pour des tâches de production langagière). En effet, le modèle TRI repose sur certains présupposés dont le plus important est l'unidimensionnalité de l'objet évalué ; or, la compétence est toujours pluridimensionnelle. Dans de nombreux tests, les concepteurs décomposent alors la tâche complexe en tâches plus simples ; ce faisant, on n'évalue plus une compétence, mais une somme de composantes (le tout n'est pas la somme des parties). On peut espérer que, d'ici quelques années, les édumétriciens auront pu faire évoluer le modèle TRI vers un modèle multidimensionnel. Une même démonstration a été faite par Behrens (2008) à partir du projet HARMOS en Suisse visant à harmoniser les objectifs de formation des 26 Ministères de l'Éducation et à concevoir une épreuve commune nationale d'évaluation, validée avec le modèle TRI. Malgré les efforts importants et un échantillon de 12 000 élèves de 6^e et 9^e, il n'a pas été possible de vérifier les propriétés métriques et les courbes de difficultés attendues par le modèle de validation TRI ; d'autre part, la stabilité du modèle à travers les différentes cultures linguistiques demeure faible à cause d'une sous-estimation de l'investissement à fournir pour tenir compte d'une situation multilingue (au plus les items sont contextualisés, au plus ils sont rejetés par le modèle ; ce qui signifie que même les applications habillées risquent de l'être).

En attendant de disposer de tels modèles et en pensant aux besoins des enseignants, nous proposons une procédure qui repose sur les principes suivants :

- Le processus de certification doit éviter les réussites et surtout les échecs abusifs : un échec est abusif lorsque l'élève ne peut passer dans une année supérieure alors qu'il était à même de suivre les apprentissages de l'année supérieure, car il maîtrisait à un seuil suffisant les critères minimaux prérequis (un échec abusif est grave en termes de conséquences, d'autant plus qu'il se décèle difficilement l'année suivante).
- L'évaluation d'une compétence suppose que l'on ait identifié les dimensions ou critères à utiliser, en distinguant les critères minimaux et les critères de perfectionnement ; l'expérience nous indique que 2 ou 3 critères et 1 ou 2 critères de perfectionnement permettent une bonne évaluation de la compétence, à condition de bien les identifier.

- Si institutionnellement l'évaluateur doit attribuer une note unique pour une épreuve d'évaluation d'une ou plusieurs compétences, deux règles combinées permettent d'éviter la plupart des échecs abusifs : la règle des 3/4 et la règle des 2/3.
 - La règle des 3/4 préconise que 3/4 de la note totale porte sur ce qui est strictement indispensable (et donc sur les critères minimaux).
 - La règle des 2/3 préconise que ce qui est fondamental puisse être évalué plusieurs fois (ce qui est le cas pour des critères de forme et de fond dans des productions langagières) et que le seuil de maîtrise de ce qui est fondamental est de 2/3 (on tolère au maximum 1/3 d'erreurs sur ce qui est strictement fondamental).
 - Si, comme dans la plupart des institutions scolaires, la note finale de 50 % est la note de passage, la combinaison de la règle des 3/4 et de la règle des 2/3 évite les échecs abusifs (car les 2/3 de 3/4 d'une note maximale est bien 50 %).

Illustrons cette procédure pragmatique pour l'évaluation d'une compétence langagière comme dans le cas d'une dissertation où l'on a identifié 3 critères minimaux et 2 critères de perfectionnement. Les critères minimaux adoptés par l'évaluateur sont les suivants :

(C1) la pertinence de la production par rapport à la tâche demandée ;

(C2) la qualité du raisonnement adopté ;

(C3) la correction linguistique (l'élève ayant à sa disposition les outils de référence).

L'évaluateur a adopté les critères de perfectionnement suivants :

(C4) l'originalité des arguments ;

(C5) le caractère équilibré de la production.

Il va de soi que chacun de ces critères aura fait l'objet d'un travail de clarification et d'un apprentissage jugé suffisant avec l'épreuve de certification. Dans le tableau suivant nous illustrons la procédure de pondération qui respecte les règles énoncées.

A nos yeux, ce qui compte avant tout, ce n'est pas la note finale, mais le degré de maîtrise de chacun des trois critères minimaux (strictement fondamentaux). Et nous avons mis en gras dans ce tableau la maîtrise minimale pour chacun des 3 critères minimaux, ce qui correspond ici à un total de 12/20, ce qui dépasse le seuil de 50 % des points. Pour être atteint, ce seuil de 50 % est surtout conditionné par les points obtenus dans les critères minimaux (15 points sur 20, c'est-à-dire 3/4 de la note maximale).

Tableau 2. L'évaluation d'une compétence langagière.

Niveaux de maîtrise	Critères minimaux			Critères de perfectionnement	
	C1	C2	C3	C4	C5
Absence totale	0	0	0	0	0
Maîtrise partielle	2	2	2	1	0
Maîtrise minimale	4	4	4	2	1
Maîtrise maximale	5	5	5	3	2
Total	15/20			5/20	

Etablie pour tenter de répondre à certaines caractéristiques souhaitables pour certifier des compétences en évitant des échecs abusifs, cette procédure est loin d'être parfaite, puisque l'obligation institutionnelle d'attribuer une note unique enfreint le principe selon lequel on ne peut en toute rigueur additionner des scores obtenus sur des dimensions ou critères théoriquement indépendants. Prôner l'adoption d'un vecteur de notes, obtenues sur chacun des 5 critères, serait déjà plus correct. La procédure adoptée ici doit être considérée comme un « moindre mal ».

5. Le renforcement nécessaire des fonctions d'orientation et de régulation

La fonction certificative est trop fréquemment utilisée dans la pratique enseignante ; elle est donc utilisée à mauvais escient, puisque en toute rigueur elle ne peut être appliquée que pour des apprentissages entièrement terminés. Par contre, les fonctions d'orientation (comme préparer une nouvelle année scolaire ou une nouvelle séquence d'apprentissage importante) et de régulation (améliorer l'apprentissage en cours) sont trop peu fréquentes, alors que le cœur du métier d'enseignant est de faire apprendre et que l'évaluation d'orientation (par l'évaluation et la consolidation des prérequis) et de régulation (diagnostiquer les difficultés et y remédier) sont des composantes clés de l'apprentissage, comme l'ont montré de nombreux travaux (voir par exemple la synthèse d'Allal & Mottier Lopez 2007). Afin que le métier d'enseignant puisse vraiment franchir le pas du professionnalisme (à comparer, par exemple, au statut du médecin ou de l'ingénieur), il est important de travailler le pouvoir diagnostique des enseignants afin qu'ils puissent prendre des décisions d'orientation des actions à entreprendre ou d'amélioration des actions en cours.

Illustrons cela à propos d'une séquence classique aussi banale que l'apprentissage de l'accord du verbe avec le sujet. Qu'observe-t-on le plus souvent de la part des enseignants ? Avec souvent beaucoup de conscience professionnelle, ils énoncent ou induisent la règle, l'illustrent, donnent quelques exercices collectifs ou individuels d'entraînement, puis font un test pour diagnostiquer les erreurs qui subsistent. Le plus souvent, ce test consiste à fournir quelques phrases à accorder, phrases inventées sur place ou puisées dans un manuel. Imaginons que la plupart des élèves aient réussi au moins 7 phrases sur 10, l'enseignant est-il suffisamment sûr de la maîtrise de la règle ? La réponse est souvent négative, car le test créé par l'enseignant n'a pas été conçu avec un raisonnement diagnostique. Il en va autrement si l'enseignant interroge son expérience professionnelle et se demande dans quels types de phrases la règle de l'accord est plus fréquemment enfreinte. Identifier des sources d'erreur et les combiner dans un tableau croisé pour générer des phrases correspondantes donne un pouvoir diagnostique bien plus puissant que des phrases conçues intuitivement ou sur le vif. Le tableau 3 illustre cette technique diagnostique : il montre qu'avec 9 phrases bien choisies, l'évaluateur pourra plus facilement prendre des décisions pour améliorer l'apprentissage en cours.

Tableau 3. Exemple de test concrétisant un raisonnement diagnostique sur l'accord du verbe avec le sujet.

	Sujet unique au singulier	Sujet unique au pluriel	Sujets multiples
Sujet + Verbe	Fati <i>corrige</i> ses devoirs	Les enfants <i>mangent</i> leur déjeuner	Ali et Fati <i>jouent</i> avec des balles
Sujet + Incise + Verbe	Ali, vêtu de beaux habits, <i>danse</i> devant ses amis	Les amis, élèves de la même école, <i>gagnent</i> le concours	Le mouton et la poule, près de la porte, <i>tentent</i> de manger un sachet
Complément + Verbe + Sujet	Dans les parcelles du jardin <i>picore</i> une poule	Sur un arbre <i>chantent</i> des oiseaux	Dans la maison nettoyée par Fati <i>entrent</i> deux étrangers

Ce tableau mérite quelques commentaires. Au départ, le test n'est évidemment pas donné de cette façon : les phrases sont données dans le désordre et le verbe est mis entre parenthèses à l'infinitif comme cela se fait classiquement. Par contre, la correction (qui peut être collective) se fera à partir du tableau croisé de telle façon que chaque élève lui-même découvre les sources d'erreur.

Notons encore que la conception des phrases mérite une attention soignée afin d'être suffisamment sûr qu'un accord correct correspond bien à une maîtrise de la règle. Ainsi, on remarquera dans le tableau que (i) toutes les phrases ont utilisé des verbes du premier groupe, car les verbes d'autres groupes peuvent induire phonologiquement l'accord, (ii) chaque phrase comprend un « obstacle épistémologique » dans le sens où, si l'accord exige le singulier, la phrase comprend également un nom au pluriel et inversement, si l'accord exige le pluriel (ex. : « Ali, vêtu de beaux habits, danse devant ses amis » ; « danse » est au singulier, mais « habits » et « amis » jouent le rôle d'obstacles). Cette technique est simple et est applicable dans de nombreux domaines d'apprentissage ; elle demande simplement que l'enseignant capitalise les résultats de ses observations et de ses analyses.

En guise de conclusion

Notre contribution a tenté de mettre en évidence la thèse selon laquelle la pratique de l'évaluation est un des révélateurs les plus puissants du professionnalisme de l'enseignant. C'est la raison du titre de cet article « Ne pas se tromper d'évaluation ». Trop fréquemment, la pratique évaluative se résume à l'attribution d'une note, alors que celle-ci à elle seule ne pourra pas remplir la fonction visée. Trop souvent, la question de la fonction n'est même pas posée : une même procédure est conduite quelle que soit la fonction qui aurait dû être visée à un moment donné de l'apprentissage. Trop souvent, on cherche à mettre une note alors qu'une démarche descriptive ou herméneutique serait bien plus pertinente. Trop souvent, le processus évaluatif consiste à émettre un jugement sur la performance de l'élève alors qu'il s'agit de prendre des décisions pour l'action. Trop rarement, les pratiques reposent sur des raisonnements diagnostiques susceptibles d'orienter une nouvelle séquence d'apprentissage ou d'améliorer un apprentissage en cours. Que d'occasions perdues de capitaliser le fruit de son expérience professionnelle.

Pourquoi ce manque de professionnalisme ? L'enseignement restera-t-il simplement un métier ou peut-il devenir une profession ? La fonction essentielle de l'enseignant (faire apprendre) n'est-il pas une fonction aussi complexe que la fonction essentielle du médecin (assurer la santé) ou de l'ingénieur (concevoir et faire fonctionner des productions) ? Si les caractéristiques essentielles de ce qu'on appelle une « profession » (Maroy 2006) sont bien (i) une activité basée sur un corps de connaissances et de techniques utilisées dans la pratique professionnelle et acquises lors d'une formation longue, (ii) une autonomie du professionnel dans l'exercice de son travail, (iii) une orientation de service à la société, (iv) une éthique de la profession, (v) une association professionnelle chargée de contrôler l'activité de ses membres et défendre la profession à l'extérieur, alors l'activité de l'enseignant correspond bien à une profession (à l'exception sans doute encore sur le dernier point). N'est-il pas temps de quitter les seuls discours théoriques pour professionnaliser l'enseignement ! Plutôt que de se contenter de faire un discours sur l'évaluation, faisons construire et discuter des dispositifs d'évaluation pour en dégager ensuite les principes qui permettront d'en construire d'autres. Et capitalisons nos expériences.

Jean-Marie De Ketele

GIRSEF (Groupe Interdisciplinaire de Recherche sur la Socialisation, l'Éducation et la Formation)

Université Catholique de Louvain à Louvain-la-Neuve

1, place de l'Université

B-1348 Louvain-la-Neuve (Belgique)

<jean-marie.deketele@uclouvain.be>

Références

- Allal, L. & Mottier Lopez, L. (éds) (2007). *Régulation des apprentissages en situation scolaire et en formation*. Bruxelles, De Boeck.
- Behrens, M. (2008). Les standards de formation, l'anneau de Moebius de l'école suisse? *Revue internationale d'éducation de Sèvres*, 48, 127-136.
- Bloom, B.S., Hastings, Th. & Madaus, G.F. (1971). *Handbook of formative and summative evaluation of student Learning*. New York, Mac Graw Hill.
- De Ketele, J.M. (2006). *Contrôles, examens et évaluation*. In J. Beillerot & N. Mosconi, *Traité des Sciences et Pratiques de l'Éducation*, Paris, Dunod, 407-419.
- De Ketele, J.M. (1996). *La formation des professeurs d'université : entre des logiques personnelles et des logiques institutionnelles*. In J. Donnay & M. Romainville, *Enseigner à l'université, un métier qui s'apprend*, Bruxelles, De Boeck, 86-112.
- De Ketele, J.M. (1993). L'évaluation conjuguée en paradigmes. *Revue française de pédagogie*, 103, 59-80.
- De Ketele, J.M. (1985). *Docimologie. Introduction aux concepts et aux pratiques*. Louvain-la-Neuve, Cabay.
- De Ketele, J.M. & Gerard, F.M. (2005). La validation des épreuves d'évaluation selon l'approche par les compétences. *Mesure et évaluation en éducation*, 28 (3), 1-26.
- De Ketele, J.M. & Roegiers, X. (1991). *Méthodologie du recueil d'informations*. Bruxelles, De Boeck.
- De Landsheere, G. (1974). *Évaluation continue et examens. Précis de docimologie*. Bruxelles-Paris, Labor-Nathan.
- Gerard, F.M. (2008). *Évaluer des compétences. Guide pratique*. Bruxelles, De Boeck.
- Gilly, M. (1980). *Maître-élève. Rôles institutionnels et représentations*. Paris, PUF.
- Jorro, A. (2007). *Évaluation et développement professionnel*. Paris, L'Harmattan.
- Maroy, C. (2006). Les évolutions du travail enseignant en France et en Europe. Facteurs de changement, incidences et résistances dans l'enseignement secondaire. *Revue française de sociologie*, 155 (2), 111-142.
- Noizet, G. & Caverni, J.P. (1978). *Psychologie de l'évaluation scolaire*. Paris, PUF.
- Pennac, D. (2007). *Chagrin d'école*. Paris, Gallimard.
- Piéron, H. (1963). *Examens et docimologie*. Paris, PUF.
- Scriven, M. (1967). *The Methodology of Evaluation*. Chicago, Rand Mac Nally.
- Stufflebeam, D.L. & al. (1980). *L'évaluation et la prise de décision en éducation*. Victoriaville (Canada), N.H.P.
- Suchaut, B. (2006). *Les évaluations nationales en France : outils pédagogiques et outils de pilotage, intérêts, limites et nouvelles perspectives*. Luxembourg, Colloque de l'ADMEE.
- Weiss, J. (1984). *La subjectivité blanchie*. Neuchâtel, IRDP.