

- Page 2 : Des évaluations omniprésentes dans le système éducatif
- Page 10 : Évaluer pour contrôler
- Page 19 : Évaluer pour former
- Page 27 : Évaluer en pratiques
- Page 35 : Conclusion
- Page 37 : Bibliographie

# ÉVALUER POUR (MIEUX) FAIRE APPRENDRE

## INTRODUCTION

Face à l'emprise des évaluations en tous genres dans l'espace éducatif, certains seraient sans doute enclins à détourner l'adage et se satisfaire d'une maxime telle que « Évaluer, évaluer, il en restera toujours quelque chose ». Les contrôles, épreuves, examens et concours de toutes sortes apparaissent tellement naturels qu'on n' imagine pas de scolarité digne de ce nom sans leur consécration. Il suffit d'évoquer les inconvénients des notes pour s'attirer rapidement la réplique outrée : « *Ce n'est pas en brisant le thermomètre qu'on résoudra les problèmes !* ».

Pourtant, cette métaphore du thermomètre est aussi fréquente qu'erronée et contre-productive. L'idée que l'évaluation intervient de façon indépendante *ex-post*, après l'enseignement, méconnaît en effet la réalité des processus scolaires. Ce que « mesurent » les évaluations, c'est en grande partie... l'effet de ces évaluations sur les apprentissages. Plutôt que de faire mine de croire que l'on mesure de façon « neutre » ce qui a été appris, les recherches convergent pour estimer qu'il serait plus pertinent de prendre en compte dès le départ qu'enseignement, apprentissage et évaluation forment un continuum.

Dès lors, l'évaluation peut devenir non plus



Par Olivier Rey et Annie Feyfant



Chargés d'étude et de recherche au service Veille et Analyses de l'Institut français de l'Éducation (IFÉ)

seulement un outil approximatif de contrôle mais aussi un levier pour mieux faire apprendre ●. C'est l'approche retenue pour ce dossier, mise à jour partielle d'un dossier publié en 2008 ●.

- Le calendrier ministériel, avec le lancement en juin 2014 de la [conférence nationale sur l'évaluation des élèves](#), coïncide avec cette préoccupation.
- Les différents types d'évaluation des élèves ont été déclinés par L. Endrizzi et O. Rey (2008).

Toutes les références bibliographiques dans ce Dossier sont accessibles sur notre [bibliographie collaborative](#).

Autant l'évaluation externe des systèmes éducatifs fait désormais l'objet de nombreux travaux (il suffit d'évoquer PISA), autant les examens nationaux (qu'on pense au baccalauréat ou au brevet) mais aussi l'évaluation au quotidien dans la classe sont étonnamment mal connus et peu étudiés ●.

À partir d'analyses, issues le plus souvent de la littérature scientifique récente, nous nous efforcerons donc de clarifier les termes du débat en revenant sur les définitions, les fonctions et les formes existantes de l'évaluation. Les pratiques de notation et les évaluations certificatives qui pèsent si lourd, comme l'évaluation « pour contrôler », seront mises en examen et comparées dans une perspective internationale. Nous nous interrogerons ensuite sur les termes d'évaluation formative et d'évaluation formatrice et envisagerons leurs implications pour les enseignants. Des dispositifs et des cadres d'évaluation qui tentent de décliner ces principes seront ensuite présentés pour contribuer à faciliter les aller-retour entre les théories présentées et les pratiques de classe existantes ou envisageables.

## DES ÉVALUATIONS OMNIPRÉSENTES DANS LE SYSTÈME ÉDUCATIF

Si l'expression de la « pression évaluative » a été utilisée par P. Merle en 2005, plusieurs ouvrages et articles récents évoquent la tension, la menace, la frénésie évaluative en même temps que la peur de l'évaluation (Butera *et al.*, 2011 ; Hadji, 2012 ; Crahay *et al.*, 2012). Le stress, la pression (contre-productive) et les différentes formes de stigmatisation amènent à constater que le système de notation utilisé en France « est une des causes du décrochage des élèves » (Merle, 2012a). Dans un rapport publié en 2013, l'Inspection générale constate que les parents trouvent leurs enfants plus confiants lorsqu'ils ne sont pas notés, même si les élèves qui réussissent aiment mesurer le plus finement possible cette réussite (Houchot *et al.*, 2013).

Chacun a pu expérimenter comment les élèves, dans nos systèmes scolaires contemporains, ont largement tendance

à investir leurs efforts en tenant fortement compte du poids donné à tel ou tel contenu par l'examen (par le jeu des coefficients) ou à quel point il est difficile d'obtenir qu'un enseignement « sans évaluation » soit pris en considération. Ce travers de la réification de l'évaluation dans l'éducation a souvent été dénoncé, puisqu'il conduit le plus souvent à mesurer ce qui est facile à évaluer plutôt qu'à évaluer ce qui a été effectivement appris (Timmis *et al.*, 2012).

## DE QUELLE ÉVALUATION PARLE-T-ON ?

Peu de processus semblent pourtant aussi « naturels » dans l'éducation que celui de l'évaluation des apprentissages, du fait de l'existence de tout un dispositif de contrôles et d'examens aux techniques toujours plus raffinées et enrichies au fur et à mesure du développement des technologies. Polysémique s'il en est, le mot « évaluation » recouvre donc de multiples processus dans le système éducatif, puisqu'il peut servir à désigner aussi bien un contrôle surveillé de l'enseignant dans sa classe, un examen national comme le baccalauréat, un examen conduisant à un diplôme, un dispositif visant à mesurer la qualité d'un établissement scolaire, voire même une procédure de contrôle des pratiques enseignantes, etc.

On peut également opérer une distinction entre l'évaluation qui est une part intégrale du processus d'enseignement et d'apprentissage et l'évaluation pour la communication, qui vise à fournir des informations à de potentiels usagers et partenaires de l'éducation, qu'il s'agisse d'étudiants, d'enseignants, d'institutions ou de systèmes (Broadfoot, 2007).

Dans le cadre de ce dossier, il n'était donc évidemment pas question d'embrasser l'ensemble du champ potentiellement couvert par l'évaluation, mais de centrer notre regard sur l'aspect particulier de l'évaluation dans et pour l'apprentissage, c'est-à-dire de l'évaluation des élèves dans le contexte particulier de situations d'enseignement de type scolaire. Encore que, même dans cette situation scolaire, on puisse distinguer plusieurs types d'évaluation significativement différents. Ainsi, il est de tradition de distinguer l'éva-

« Bien qu'il s'agisse d'un rituel quasi-quotidien de la classe, les évaluations constituent encore des dispositifs peu étudiés en didactique » (Millon-Fauré, 2013).

Science dont le nom est attribué à Henri Piéron et dont le but est l'étude de l'organisation des examens, de leurs contenus, objectifs pédagogiques mais aussi (et surtout) des méthodes de correction des épreuves ainsi que le comportement des examinateurs et des examinés.

luation selon les fonctions qu'elle remplit dans l'apprentissage : diagnostique, pronostique, sommative, certificative, formative...

Les « contrôles » ou « examens » de type sommatif rencontrés tout au long de la scolarité, auxquels on pense le plus spontanément, ne sont pas de même nature que les évaluations certificatives (concours diplômants par exemple), les évaluations pronostiques (examen d'accès ou d'orientation) ni les évaluations diagnostiques (grandes enquêtes par niveau de scolarité), sans enjeu « direct » pour l'évalué. Quant à l'évaluation formative, elle est souvent mêlée aux autres formes d'évaluation, dans la mesure où c'est son utilisation dans l'apprentissage qui la distingue.

Toutes ces différentes formes d'évaluation n'ont pas les mêmes effets sur les personnes, au-delà même du contexte scolaire puisque l'évaluation sommative, par exemple, devient fréquemment un marqueur de l'identité scolaire de l'élève, du fait qu'elle est souvent enregistrée dans un document officiel et public, tel que le bulletin scolaire (Allal, in van Zanten, 2008).

Si la certification a longtemps constitué le seul débouché de l'évaluation, la démocratisation scolaire qui s'est développée depuis les années 1960 a mis en avant le souci d'évaluation comme processus de vérification continue pour guider la démarche d'enseignement et d'apprentissage (Scallon, 2007). Ce qui explique que les réflexions ne se réduisent plus aujourd'hui à une question d'exactitude de la mesure ni de produit fini (résultat du test par exemple) mais portent également sur les progressions des élèves.

On le constate aussi dans le vocabulaire utilisé dans la recherche de langue anglaise, dans laquelle les termes de *measure*, voire de *testing*, ont quasiment disparu, pendant que celui d'*evaluation* est désormais plutôt réservé aux organisations ou aux systèmes, pour céder la place à *formative assessment* ou *classroom assessment*.

La recherche de procédés fidèles et valides, exempts de subjectivité, qui jugent à un moment précis et de façon isolée tout ce qui est censé constituer l'acquis d'un individu, n'est plus la priorité, même si cela a constitué pendant une période l'horizon de la docimologie ●.

Évaluer est toujours un jugement en fonction d'une valeur, et l'enjeu n'est donc pas tant de rendre l'évaluation plus exacte et plus juste, mais plutôt de communiquer à l'évalué ce qu'on attend de lui et de l'inciter ainsi à partager les finalités de la formation.

Si cette mesure « objective » est pourtant bien ce qui est souvent encore cherché au travers des pratiques de notation qui constituent la caractéristique majeure des dispositifs d'évaluation existants, la plupart des experts défendent l'idée que l'évaluation est un « message » plus qu'une « mesure » (voir aussi le [dossier en ligne](#) réalisé par Jacques Nimier : « Cette évaluation impossible et pourtant nécessaire »).

## ÉVALUER LES ÉLÈVES OU ÉVALUER LE SYSTÈME ?

La question de l'évaluation des élèves est intimement connectée à celle de l'évaluation du système éducatif. Que l'on considère les classements des lycées en fonction des résultats du bac, les évaluations diagnostiques réalisées par la Direction de l'évaluation, de la prospective et de la performance du ministère de l'Éducation nationale, ou encore les enquêtes internationales PISA (Rey, 2011), on comprend qu'à travers les résultats des élèves, c'est souvent la performance des enseignants, des approches pédagogiques, des établissements ou des systèmes éducatifs qu'on cherche à mesurer. Elles renvoient à l'effet-miroir cher à Claude Thélot, qui permettrait notamment aux enseignants de prendre de la distance par rapport à leur propre pratique en comparant « leurs » résultats par rapport à ceux des autres placés dans des conditions comparables.



## Évaluation n'est pas forcément comparaison

Plusieurs chercheurs s'interrogent sur la frénésie évaluative qui s'est emparée de la société. L'évaluation est très souvent source de comparaison : comparaison internationale avec les évaluations PISA ; comparaison entre écoles, avec la publication des rapports d'inspection en Angleterre, etc.

Black et Wiliam (1998) n'ont cessé de dénoncer les difficultés posées par une évaluation basée sur des tests, notes et comparaisons. L'esprit de compétition influe sur tous les acteurs de cette compétition : les élèves vont fournir des efforts pour le test et au moment du test ; les enseignants vont encourager la mémorisation et un apprentissage superficiel (mais rentable). La polarisation sur les tests nationaux ou internationaux induit des apprentissages et un enseignement *a minima* (pour afficher un taux de réussite correct) ●. Les stratégies pédagogiques s'en trouvent détournées au détriment des besoins d'apprentissage des élèves, les objectifs cognitifs sont survalorisés au détriment d'objectifs de socialisation plus larges.

La montée en puissance, depuis les années 1980, d'une demande générale pour que le système éducatif puisse « rendre des comptes » publiquement, régulièrement et à tous les niveaux, sur ses performances (*accountability*), a en effet renforcé cette pression à la production d'indicateurs standardisés, par des examens ou des tests à grande échelle, et la polarisation proportionnelle du système éducatif sur ces épreuves sommatives qui concentrent de nombreux enjeux critiques (*high stakes testing*). Dans la plupart des pays de l'OCDE, de grandes épreuves ont été mises en place pour traduire ces politiques d'évaluations standardisées, malgré des résultats douteux en termes d'efficacité comme en termes d'égalité (Mons, 2009).

Deux auteurs britanniques proposent, de ce point de vue, une lecture des réformes de l'école primaire anglaise à partir d'un bilan de plusieurs recherches (Broadfoot & Pollard, 2006). Selon eux, *via* l'imposi-

tion de nouvelles pratiques d'évaluation, le modèle de la performance a progressivement envahi les écoles primaires mettant un terme à leurs traditions précédentes d'éducation libérale et progressive. À l'évaluation basée sur le développement personnel et des critères diffus, le modèle de la performance a substitué une évaluation basée sur des productions et des critères explicites et chiffrés (*levels, targets, standards, league tables, value added, etc.*). Aux yeux des chercheurs, ce modèle a favorisé une tension croissante dans les établissements scolaires, le climat de compétition et de concurrence allant de pair avec un accroissement de la violence et de la polarisation sociale sur fond de moindre solidarité. Ces effets du modèle de la performance peuvent s'avérer finalement contre-productifs à terme, y compris sur un plan économique, car contradictoires avec l'idéologie d'une société de la connaissance tout au long de la vie, dans la mesure où la satisfaction d'indicateurs de court terme décourage l'apprentissage à plus long terme, particulièrement pour ceux qui échouent dans leur scolarité initiale.

Finalement, en matière de politique éducative, le mauvais alignement entre les évaluations en classe et les évaluations standardisées externes à forts enjeux sont susceptibles de saper les efforts déployés par les enseignants en classe (Looney, 2011). Dans la plupart des pays, les formes d'évaluation restent assez traditionnelles, quelles que soient les évolutions des contenus d'enseignement. Elles tendent à concentrer l'attention des acteurs éducatifs sur un nombre limité de résultats cognitifs : le curriculum implicite mais réel est alors singulièrement plus étroit que le curriculum officiel (Nusche *et al.*, 2013).

Les technocrates ont « donné de la valeur à ce qu'ils mesuraient au lieu de mesurer ce à quoi ils donnaient de la valeur » (Hargreaves & Shirley, 2009, cités par Muller & Normand, 2013).

Pour l'OCDE (Nusche *et al.*, 2013), des priorités communes à tous les pays peuvent être définies en matière d'évaluation :

- les différents composants de l'évaluation doivent former un tout cohérent pour être efficaces dans l'amélioration des performances éducatives ;
- l'évaluation doit être alignée par rapport aux objectifs éducatifs globaux, qui doivent être clairement compris par les acteurs de l'éducation ;
- tous les types d'évaluation devraient avoir une valeur éducative (en sus de leur rôle éventuel de contrôle), c'est-à-dire entraîner un clair bénéfice pour les enseignants et les élèves ;
- il convient d'éviter les distorsions dues aux évaluations standardisées et à l'utilisation des résultats pour rendre compte de l'efficacité du système (accountability), l'un des moyens étant par exemple de veiller à la diversité des approches pour mesurer la performance éducative ;
- la mesure des performances doit être large, à la fois qualitative et quantitative, et centrée sur le développement de l'élève et ses progrès.

### Des enquêtes PISA essentiellement lues à l'aune de la compétitivité

On peut remarquer que le terme de performance est largement utilisé pour rendre compte des résultats des enquêtes PISA, qui questionnent des jeunes de 15 ans, alors même que le développé du sigle signifie « Programme international pour le suivi des acquis des élèves ». En France, la direction de l'évaluation, de la prospective et de la performance (DEPP) « *exerce ses compétences d'évaluation et de mesure de la performance dans les domaines de l'éducation et de la formation. Elle contribue à*

*l'évaluation des politiques conduites par le ministère de l'Éducation nationale* ». Performance et évaluation sont associées pour évoquer un outil permettant de piloter l'institution par les résultats, de piloter l'établissement, d'aider les enseignants dans leurs pratiques et les familles dans leurs attentes.

La notion de compétitivité n'est jamais très loin quand sont évoqués ces dispositifs. La [notice consacrée à PISA dans Wikipedia](#) utilise le terme de performance (des systèmes éducatifs) et évoque les disparités entre systèmes et entre établissements, ce qui témoigne d'une vision courante du programme de l'OCDE.

Ces objectifs induisent un emboîtement des évaluations, rapidement inextricable pour qui voudrait modifier le système de référence : l'évaluation des élèves sert à l'évaluation des enseignants (et/ou des écoles) qui sert à l'évaluation du système (Hadji, 2012).

La fiabilité et la validité attachées à ce type d'évaluation quantitative, selon les critères psychométriques habituels, amènent souvent à oublier que cette approche est normative plutôt que critériée : le résultat de l'élève est comparé à un groupe plutôt qu'à un niveau attendu. Ce type d'évaluation contient donc en lui-même un effet de classement, même s'il peut être canalisé (anonymat, publicité, etc.), effet peu compatible avec une évaluation des compétences qui impliquerait plutôt une approche critériée (Marcoux *et al.*, 2014).

On constate donc que même dans le cas d'évaluations qui ne sont pas directement liées à une certification ou des allocations de moyens, la forme de l'évaluation induit un regard déformant sur les acquis des apprentissages et contribue en retour à en légitimer certains plus que d'autres, sans que l'on en ait une conscience très claire.

### Des évaluations nationales pour améliorer l'enseignement ?

Qualifiées de « dispositifs d'évaluation pédagogique » lors de leur mise en place, en France, à la fin des années 1970, les évaluations nationales n'ont pourtant rien de formatif (Hadji, 2012). Initiées en 1979 pour créer un « observatoire permanent », au niveau du cours moyen



deuxième année (CM2) et de la 6<sup>e</sup> ou du cours préparatoire (CP), ces évaluations ont concerné, selon les années, les élèves du primaire ou du collège, voire de la seconde (en 1989). L'objectif de ces évaluations était de connaître le degré de réalisation des objectifs officiels. À la fin des années 1990, les objectifs affichent une volonté plus pédagogique, les évaluations de fin de primaire deviennent un « *outil pédagogique pour l'enseignant pour définir ses actions de remédiation* » ●. Elles restent un élément de comparaison sur une décennie, par exemple, mais s'accompagnent d'analyses plus poussées sur les facteurs influant les performances et les acquisitions des élèves, pour une année donnée et par niveau (CE2 et 6<sup>e</sup>). À partir de 2001, la direction de la programmation et du développement qualifie l'enquête annuelle (3 000 élèves par an) non pas comme un bilan de l'école primaire mais comme une « *mise en avant des compétences acquises ou en cours d'acquisition sur lesquelles l'enseignant pourra s'appuyer pour élaborer ses programmes* » (Note d'information, n° 01.59).

À qui sont destinées ces évaluations de masse ? Essentiellement au grand public et aux enseignants. Pour le grand public, il s'agit de faire le point sur l'état des savoirs et savoir-faire des élèves sur certaines compétences. Les compétences sont classées en trois grandes catégories : compétences de base, compétences approfondies et compétences remarquables. L'évaluation va indiquer si l'élève ne maîtrise pas les compétences de base, s'il maîtrise uniquement celles-ci, s'il maîtrise aussi les compétences approfondies et s'il maîtrise en outre les compétences remarquables. En 1998, on présente les scores moyens par champ en français et en mathématiques, et on parle de « profil des élèves ». En 2001, on détecte les « points sensibles ».

Quelles sont les limites de ces évaluations « de masse » ? Il n'y a pas véritablement de comparaison de performance dans le temps, les échantillons sont de fait différents mais surtout les épreuves sont différentes.

## Les évaluations nationales ailleurs, entre contrôle et diagnostic

Ailleurs, d'autres formes d'évaluations nationales interpellent aussi sur l'utilisation qui peut en être faite pour améliorer l'enseignement. Ainsi, aux **États-Unis**, la comparaison de cohortes d'élèves de primaire sur quatre ans, au regard de leurs résultats aux évaluations standardisées en lecture, a montré que les tests basés sur les standards ne procuraient que des informations trop pauvres pour établir un diagnostic utile et masquaient une hétérogénéité significative des causes de médiocres performances derrière des résultats équivalents (Rupp & Lesaux, 2006).

[Un exemple d'évaluations standardisées et leurs écueils : en Pennsylvanie](#)



En Belgique, la scolarité des élèves de la **fédération Wallonie-Bruxelles (FWB)** est rythmée par plusieurs évaluations externes non certificatives organisées chaque année pour les élèves inscrits en 3<sup>e</sup> et 5<sup>e</sup> primaire et pour ceux de 3<sup>e</sup>, 4<sup>e</sup> ou de 5<sup>e</sup> année de l'enseignement secondaire. Elles portent successivement sur la lecture et la production d'écrit, sur les mathématiques et sur les sciences et l'éveil. Une extension aux langues vivantes est à l'étude. Les résultats de ces évaluations n'affectent pas le parcours scolaire des élèves ; ils sont mis à la disposition des chefs d'établissement et des enseignants à titre diagnostique. À chaque session, le Service général du pilotage du système éducatif de la FWB produit des pistes didactiques en vue d'aider les équipes enseignantes à améliorer la maîtrise des compétences ciblées et à développer une pédagogie différenciée, conformément aux prescriptions du décret « Missions » (Endrizzi, 2013).

Au **Danemark**, au tournant des années 1980, les inévitables résultats PISA (jugés décevants) ainsi qu'une étude de l'OCDE poussent le gouvernement à prendre des mesures pour améliorer la culture de l'évaluation dans le système. Des tests nationaux sont mis en place en 2008 pour suivre les acquis des élèves au cours de

● Chaque année en 1997, 1998 ou 1999, les *Notes d'information* du ministère de l'Éducation nationale dressent le tableau des profils et compétences des élèves en mathématiques.

L'agence Skolverket met également à disposition des enseignants une banque nationale de tests dans plusieurs matières et travaille au déploiement d'une plateforme pour l'évaluation en ligne.

De Ketele parle aussi de fonction de régulation.

la scolarité dans les grands domaines (danois, anglais, mathématiques, géographie, biologie, physique-chimie). Les résultats restent confidentiels. Ils ne sont publics qu'à leur niveau national agrégé et ne sont communiqués qu'à la famille, sans les résultats au niveau de la classe ou de l'école, pour éviter toute dérive en termes de palmarès. L'objectif est de n'utiliser les tests que comme un outil pédagogique pour les enseignants, en vue d'adapter leurs pratiques, si nécessaire (Lassen, 2013).

En **Suède**, les évaluations externes, organisées par Skolverket, rythment la scolarité à la *grundskola* (en 3<sup>e</sup>, 6<sup>e</sup> et 9<sup>e</sup> années), tandis qu'à la *gymnasieskola*, elles ciblent chaque année les élèves qui ont terminé leurs cours de suédois, de mathématiques et d'anglais. Le gouvernement étudie actuellement la possibilité d'étendre les épreuves à d'autres matières, pour favoriser des pratiques d'évaluation plus homogènes et assurer un meilleur suivi des élèves ● (Endrizzi, 2013).

## DES CONFUSIONS DANS LES USAGES

« Une évaluation est valide si elle contient des informations rendant possibles des apprentissages ultérieurs et si les enseignants les utilisent pour ajuster leur enseignement » (Rémond, 2008).

### Des fonctions à préciser

Comme l'explique De Ketele, on peut distinguer trois fonctions différentes de l'évaluation, qui préparent toutes une prise de décision : la fonction certificative (échec ou réussite), la fonction formative (pour améliorer les apprentissages) ● et la fonction d'orientation (pour préparer une nouvelle action). On constate bien souvent que les pratiques privilégiées sont certificatives. Pour De Ketele (2013), cela tiendrait à la confusion entre la fonction de l'évaluation (pour quoi ?) et la démarche évaluative (comment ?). Il distingue, là encore, trois démarches : sommative (faire la somme de points acquis), descriptive (identification et description des acquis et des processus positifs) et interprétative (« donner du sens à un ensemble d'indices quantitatifs ou qualitatifs »).

Pour lui, la notation (ou plus généralement la démarche sommative) ne devrait pas concerner des apprentissages en cours d'acquisition sauf à produire de l'échec ou de la réussite abusifs. Le processus d'évaluation consiste à recueillir un ensemble d'informations pertinentes à confronter, par une démarche adéquate, à un ensemble de critères cohérents avec le référentiel, valides dans l'opérationnalisation, fiables dans l'utilisation. Les résultats de cette confrontation doivent permettre une prise de décision cohérente avec la fonction visée par l'évaluation (il s'agit de donner du sens à l'évaluation, aux regards des objectifs d'apprentissage, par exemple).

### Qu'est-ce qu'on évalue vraiment ?

On a vu que de nombreuses confusions sont souvent opérées entre les méthodes d'évaluation qu'on met en œuvre à grande échelle pour mesurer les acquis des élèves et les évaluations pratiquées quotidiennement par les enseignants. Or, « il apparaît illusoire de reproduire à l'échelle de la classe des démarches de construction et de validation d'instruments comparables à ceux employés dans des recherches évaluatives au niveau du système » (Tessaro, 2013). Autrement dit, tenter de reproduire PISA dans son école n'est peut-être pas une piste très féconde.

On estime aussi de façon implicite que l'évaluation scolaire n'évalue que les apprentissages dispensés explicitement par les enseignants : l'évaluation ne porterait donc que sur ce qui a été enseigné : discipline et programme effectif. Pourtant, la recherche a montré qu'évaluer signifie bien souvent aussi évaluer autre chose, que l'on désigne parfois par le concept de « curriculum caché », pour qualifier les savoirs et habiletés implicites liées à une maîtrise souvent familiale de la culture scolaire (« le métier d'élève »), rarement enseignée en tant que telle à l'école. La sociologie du curriculum a montré de façon assez robuste comment l'appropriation des connaissances scolaires qu'on évalue « à la fin » est largement produite par une incorporation de règles et de représentations, qui sont faiblement explicitées tout au long de la scolarité (Forquin, 2008).



Par ailleurs, on peut se demander dans quelle mesure il faut prendre en compte l'ensemble des savoirs, compétences et habiletés acquises par les élèves, à l'occasion de la vie scolaire (voyages scolaires, stages, débat sur la note de vie scolaire, participation aux activités associatives dans l'établissement, aux instances de représentation, etc.) ou à côté de la vie scolaire (engagement citoyen, activités bénévoles, projets extrascolaires ou parascolaires, etc.).

Le développement des expériences de webfolio en France constituent un pas vers la prise en compte de ce qui est maîtrisé par la personne, au-delà du cadre et des seules disciplines scolaires (voir par exemple [l'expérience du lycée pilote innovant international](#) de Poitiers). Le webfolio à la française est néanmoins essentiellement conçu, pour l'instant, dans une perspective d'aide à l'orientation et au projet professionnel.

### Un sujet fréquent de malentendus entre élèves et enseignants

Chacun a intégré plus ou moins consciemment que l'évaluation déborde largement sa fonction de « diagnostic » ou de « certification » pour imprégner l'ensemble de la vie pédagogique.

Il suffit par exemple de constater que les élèves ajustent souvent leurs efforts à la présence (ou non) d'évaluations et orientent leur attention en fonction des épreuves finales. Cet « effet en retour » des évaluations sommatives (joliment appelé *backwash effect* en anglais) est bien connu, mais le malentendu perdure : on fait toujours comme si l'évaluation arrivait de façon « autonome » après l'apprentissage, alors qu'elle façonne en large part ce dernier.

Pour beaucoup d'enseignants, les évaluations sont aussi un outil nécessaire pour motiver ou « tenir » leur classe. En avant-propos d'un ouvrage récent sur l'évaluation, une enseignante du Luxembourg n'hésite ainsi pas à affirmer que « *la plupart des profs sont profondément convaincus que les élèves apprennent uniquement sous la menace de mauvaises notes, d'ajournement*

*et de redoublements* » (Serres, 2014). Pourtant, Barrère (2014) constate que de nombreux enseignants vivent l'évaluation comme une corvée, symbolisée par les paquets de copie, et comme une zone d'activité à contingerter pour qu'elle ne prenne pas le pas sur les autres activités d'enseignement. Confrontés au dilemme d'être entraîneur ou arbitre, les enseignants ne vivent pas toujours très bien leur rôle d'évaluateur (Merle, 1998).

Ils reprochent aux élèves de donner trop de poids aux évaluations... reproche que les élèves leur retournent volontiers ! Pour ces derniers, l'évaluation, loin d'être une seule routine technique, est toujours une épreuve qui met en jeu la reconnaissance de la personne, même quand ils affectent une certaine désinvolture ou revendiquent leur stigmatisation par les résultats dans une sorte de « renversement héroïque » (Barrère, 2009). Ainsi, une « mauvaise » note ou une évaluation ratée peut rapidement générer une blessure affective, car l'évaluation de la personne est souvent, à tort ou à raison, perçue derrière l'évaluation du travail scolaire.

En outre, le « contrat » qui préside aux normes d'évaluation est souvent opaque aux yeux des élèves : il n'est pas toujours évident de comprendre par exemple que l'injonction à l'expression personnelle (« ne pas répéter bêtement ») ne signifie pas donner « son opinion », ou que l'existence de critères pour juger de la qualité d'un travail ne revient pas à additionner les points des différentes parties pour calculer la note finale.

Certains élèves ne comprennent pas non plus que la pression scolaire récurrente ne sanctionne pas que la somme de travail mise en jeu (« à travail égal, note égale »). Ainsi, des résultats d'évaluation médiocres sont douloureusement ressentis quand ils sanctionnent un investissement personnel intense.



## L'ÉVALUATION SOUS L'EMPRISE DE LA MESURE

Les théories de l'évaluation ont d'abord longtemps été caractérisées par la recherche de la « bonne » mesure, même si actuellement, cette notion de mesure absolue semble moins centrale dans le processus d'évaluation.

### Des efforts de modélisation...

Les tentatives de modélisation de l'évaluation ne manquent pas. Divers cadres théoriques ont été proposés pour tenter une adéquation entre analyse des tâches d'apprentissage et définition des objectifs. Il faut resituer les premières analyses du début du XX<sup>e</sup> siècle dans le contexte de l'époque, alors que l'évaluation ne s'entendait que comme sommative ou certificative. Évaluer consistait à mesurer. Dès la fin des années 1930, le constat d'un écart important entre correcteurs des épreuves du baccalauréat a motivé la recherche de mesures plus équitables.

En simplifiant grandement, on peut tout d'abord évoquer l'approche psychométrique (Binet, Cronbach) qui cherche à mesurer les performances des élèves par rapport à une moyenne générale (mesure relative) ou par rapport à l'écart au seuil de maîtrise (mesure absolue).

L'approche éduométrique va tenter ensuite d'adapter la théorie de la généralisabilité de Cronbach, en prenant en compte d'autres facteurs (les individus, les questions posées, les phases d'apprentissage) et en introduisant la notion de progrès et de marge d'erreur. Pour autant, cette approche reste très statistique et ne permet toujours pas de prendre en compte la diversité des élèves, d'une part, et de contrôler tous les paramètres, d'autre part.

Pour Cardinet, les mesures en psychologie ne sont pas adaptables en l'état, en pédagogie, du fait de l'absence de caractéristiques stables, d'une multitude d'objectifs, de démarches, de situations, rendant une standardisation inacceptable. L'articulation individu-situations-tâche est difficilement contrôlable et quantifiable : « *la situation joue le rôle de médiateur entre le sujet et la tâche* » ; l'élève évalué

évolue de par la phase d'évaluation, etc. (Cardinet, 1989). La théorie de la généralisabilité distingue donc quatre différenciations : des élèves (mesure de leurs différences) ; des objectifs et domaines d'enseignement ; des conditions d'apprentissage ; des niveaux d'apprentissage.

Rapprochant évaluation formative et approche par compétences, Scallon (2007) propose de construire une grille permettant d'évaluer le degré d'acquisition des compétences. Composée « d'échelles descriptives » plus ou moins nombreuses, plus ou moins détaillées, cette grille suppose de définir clairement les caractéristiques des compétences et productions attendues puis d'associer à chacun de ces éléments descriptifs un niveau de qualité correspondant à des étapes ou niveaux de progression. On verra plus loin en quoi ces éléments apparemment normatifs s'inscrivent dans une stratégie d'évaluation pour les apprentissages.

### ... à la critique de la mesure

Contrairement à ce que pouvaient en penser les premiers évaluateurs et les premiers docimologues, mesurer n'est pas synonyme d'évaluer. « *La mesure peut être un instrument efficace pour traiter une information nombreuse, devant être agrégée, et comparée sur un large territoire par exemple* » (Mottier Lopez, 2013). Mais l'évaluation des apprentissages ne peut se réduire à un objet technique et statistique. Dans les faits, l'évaluation a-t-elle une valeur intrinsèque ou extrinsèque ? Que signifie vraiment le point vert (acquis), orange (en cours d'acquisition), rouge (non acquis) ou un 12/20 dans telle ou telle matière ? Atteindre la moyenne signifie-t-il maîtriser l'essentiel ou être sur le chemin de sa maîtrise ? Que signifie une moyenne qui agglomère des informations et des compétences différentes ? Un 14/20 en mathématiques peut-il « compenser » un 6/20 en français ou en EPS ?

Cet usage de la mesure, signe d'une « quantophrénie » de la société (Hadji, 2012 ; Del Rey, 2013), est synonyme de contrôle, de sanction, mais aussi de solution de facilité tant il est difficile d'évaluer « l'humain » et plus confortable de se référer à « des procédures, méthodologies, instruments, dispositions institutionnelles



*propres à rendre l'évaluation aussi objective que possible* » (Rey, 2012), là où certains y verraient de l'arbitraire.

On constate aujourd'hui en France, dans les textes officiels, la prise en compte de certaines critiques et la recherche d'une évaluation qui ne se cantonne pas à la fonction de mesure *ex-post* des apprentissages. « *Dans l'enseignement primaire, l'évaluation sert à mesurer la progression de l'acquisition des compétences et des connaissances de chaque élève. Cette logique d'évaluation est aussi encouragée dans l'enseignement secondaire* » (Loi d'orientation et de programmation pour la refondation de l'école de la République, 8 juillet 2013, art. 34).

Dans l'annexe de la loi de programmation de 2013, un paragraphe est consacré à l'évaluation et à la nécessité de faire évoluer les modalités d'évaluation et de notation des élèves : « *pour éviter une "notation-sanction" à faible valeur pédagogique et privilégier une évaluation positive, simple et lisible, valorisant les progrès, encourageant les initiatives et compréhensible par les familles. En tout état de cause, l'évaluation doit permettre de mesurer le degré d'acquisition des connaissances et des compétences ainsi que la progression de l'élève. Il faut aussi remédier à la difficulté pour les enseignants d'évaluer les élèves avec des dispositifs lourds et peu coordonnés entre eux. Ainsi, l'évolution des modalités de notation passe notamment par une réforme du livret personnel de compétences actuel, qui est trop complexe, et une diversification des modalités de l'évaluation* ».

Actuellement, la note est encore centrale dans les dispositifs d'évaluation et les différentes parties prenantes du système éducatif attachent toujours une certaine importance à la fonction de contrôle qu'ils assument en fin de compte aux évaluations.

## **ÉVALUER POUR CONTRÔLER**

« *Les pratiques d'évaluation s'inscrivent toujours dans un contexte socio/économico/politique qui leur donne une coloration propre* » (Hadji, 2012).

## **COMMENT EN EST-ON ARRIVÉ À L'« ÉVALUATION TRIAGE » ?**

L'instruction est pendant longtemps restée relativement éloignée des préoccupations de l'évaluation normée. En fait, jusqu'à la fin du Moyen-Âge, le modèle dominant était le modèle humaniste très lié au préceptorat, pour lequel le contrôle et la mesure n'étaient pas une nécessité. Seule une élite restreinte avait alors accès à l'instruction.

Avec l'ouverture relative de l'instruction au-delà de la noblesse et du clergé, les jésuites, au XVI<sup>e</sup> siècle, imaginent une évaluation notée et une uniformisation éducative, avec un objectif essentiellement idéologique (conquérir les esprits). L'évaluation est alors plus destinée à forger des âmes catholiques qu'à connaître le niveau de chacun.

C'est le passage au modèle collectif de la classe à partir de la fin du XVII<sup>e</sup> siècle, qui amène avec lui le besoin de domestiquer le corps, de le discipliner, d'assurer une surveillance dans le cadre d'une transmission de masse selon un modèle qui introduit l'idée d'honnête émulation entre élèves. « *Ce qui caractérise la pédagogie des Jésuites, c'est la discipline, la répétition et la concurrence perpétuelle entre les élèves* » (Maulini, 2003).

Parallèlement, le modèle élitiste se déploie sous l'influence du mandarinat chinois, marqué par la méritocratie par concours pour la sélection des administrateurs de l'Empire. Le classement devient une donnée essentielle pour juger un élève au sein d'une hiérarchie donnée (la classe, le groupe, l'établissement, etc.).

La création de l'École des Ponts et Chaussées à la fin du XVIII<sup>e</sup> siècle imprime durablement sa marque en France, en consacrant l'idée de la formation d'une élite très performante capable de diriger tout un pays, dont les principes de classement et de notes vont se diffuser pour former le modèle méritocratique du système académique (Kerrero, 2014).

Au XIX<sup>e</sup> siècle, l'influence des théories scientifiques puis psychologiques conforte l'idée qu'il est possible « *d'éta-*

blir une échelle numérique sur laquelle pourront être positionnés les phénomènes observés afin de les appréhender de manière rationnelle » (Demeuse et al., cités par Mottier Lopez, 2013). Si en France la République s'oppose formellement à l'éducation religieuse à partir du XIX<sup>e</sup> siècle, c'est pour mieux récupérer certains de ses principes pour son propre compte, dont notamment cette centralité de la répartition des étudiants, assise sur des chiffres censés représenter une objectivation du mérite de chacun. Le système éducatif se centralise et s'organise en niveaux, degrés successifs, programmes spécifiques, examens par niveau, le début de siècle étant marqué par la création du baccalauréat (1808) et la fin par celle du certificat d'études primaires (1882). La recherche d'une rigueur scientifique de la notation puis d'une notation objective (psychologie scientifique), avec le développement de la docimologie (début du XX<sup>e</sup> siècle), évolue avec la massification et la « démocratisation » de l'école vers celle d'une notation équitable, qui permettrait une égalité des chances.

L'utilité de la notation (forme universelle et omniprésente de l'évaluation des élèves) répond à un projet pédagogique de connaissance (aide au développement de l'élève) mais aussi à un projet sociopolitique, répondant à un besoin social de contrôle (Hadji, 2012). Après les âmes, le but de l'école a été de trier les élèves pour les faire rentrer dans les bonnes cases, les élites d'un côté, les bons techniciens de l'autre, des paysans ne parlant plus le patois, etc. « L'école est une sorte d'immense gare de triage et d'étiquetage des personnes selon leurs compétences. La société du mérite est donc inévitablement une société de la mesure » (Merle, 2007).

En 1968, lors du fameux [colloque d'Amiens](#), le ministre A. Peyrefitte propose que soit mis fin au système de notation chiffrée pour s'intéresser et mesurer les progrès individuels selon une échelle de A à E. Son successeur E. Faure confirme l'objectif en 1969, mais l'opposition des enseignants et une certaine frilosité institutionnelle entraîne le rétablissement, en 1971, avec le ministre O. Guichard, de la note sur 20.

## NOTATION ET ÉVALUATION : UN COUPLE QUI FAIT TROP BON MÉNAGE

Pour paraphraser Jaurès, l'évaluation semble le plus souvent porter en elle la note comme la nuée porte l'orage. Incontournable dans le domaine de l'évaluation scolaire, la notation apparaît pourtant comme allant de moins en moins de soi dès lors que l'on examine de près son utilisation. En effet, de nombreux travaux mettent en lumière à la fois ses effets contre-productifs en termes d'apprentissage mais également sa fiabilité très relative pour mesurer les acquis scolaires. Ce constat n'est pas une caractéristique de la recherche en éducation française. De la Suède à l'Australie, en passant par les États-Unis où la Suisse, la notation chiffrée n'a pas fait la preuve de son efficacité quant à l'acquisition des apprentissages. Son usage a été abandonné, détourné, adapté mais rarement glorifié.

« Âgées d'un siècle et demi, les notes apparaissent aujourd'hui encore comme le symbole même de la méritocratie. Dans notre imaginaire, ces notes sentent bon la plume sergent et l'encre violette. Elles symbolisent l'école laïque et républicaine, celle qui récompense chacun selon ses mérites, et pourtant... » (Radier, 2014).

« La notion d'unidimensionnalité est une trahison de la compréhension du monde. Dès que l'on ramène une chose à un chiffre, il n'en reste plus rien. Un caillou ne vaut pas 10. Il est grand, petit, lourd, il est dur, mais il ne vaut pas 10. De la même manière dire d'une copie qu'elle vaut 15 est une stupidité. [...] la copie a un profil, elle est bonne pour les idées, mauvaise pour l'orthographe, etc. [...] La seule justification de l'unidimensionnalisation, c'est de hiérarchiser » (Jacquard, 2000, cité par Castincaud & Zakhartchouk, 2014).



## Un horizon indépassable ?

Les notes concentrent de nombreuses attentes dans le système scolaire et apparaissent encore comme la figure par excellence de l'évaluation, nécessairement sommative, dans la mesure où l'évaluation formative apparaît encore, aux yeux du plus grand nombre, comme un complément facultatif plus que comme un autre type d'évaluation possible.

La notation est universelle, anciennement ancrée dans les pratiques du système éducatif, et plus encore en France, où aussi bien les élèves que les enseignants sont évalués par des notes. La justification de cette notation est souvent facteur de dérive, biais et iniquité de la part des acteurs impliqués dans ce type d'évaluation.

La note semble être une obligation incontournable du métier d'enseignant : pour le passage dans la classe supérieure ; lors d'un devoir sur table, pour évaluer un apprentissage et rendre une note ; par une contrainte administrative (Merle, 2007, 2012b).

Certains y voient également la marque d'un pouvoir : « *Chaque enseignant détient un pouvoir sur un territoire pédagogique qui lui est considéré comme personnel : il enseigne, il crée des épreuves dont il définit les barèmes, il corrige, il note, il décide des coefficients de ces épreuves, il fait la moyenne...* » (Desclaux, 2014). Les notes sont aussi un outil essentiel pour asseoir l'autorité pédagogique de l'enseignant (ou sa popularité), ainsi que pour sanctionner ou gratifier des comportements en classe, au-delà de la vérification d'un acquis cognitif, voire pour attester d'une progression tout au long de l'année qui « valide » en quelque sorte le travail enseignant effectué.

Si la note leur donne du pouvoir, les enseignants la considèrent également comme un instrument de motivation extrinsèque. L'enseignant participe à l'opération de triage des élèves dès lors qu'il produit une note non pour évaluer les acquis des élèves, ajuster sa pratique ou les apprentissages, mais pour justifier l'appréciation donnée en conseil de classe (Desclaux, 2014). Les enseignants (ou l'établisse-

ment scolaire) subissent aussi la pression parentale, puisqu'on constate, de la part des parents, une défiance envers une évaluation non notée, souvent due au sentiment d'être dépassés par la signification des formulations employées sur le livret. Autre élément favorable aux notes, le besoin des parents de connaître le classement de leurs enfants par rapport aux autres élèves.

Au Québec, la question de la présence ou non d'évaluations chiffrées a fait l'objet de fortes polémiques qui ont parfois semblé occuper l'essentiel du débat sur la mise en place du Renouveau pédagogique : une partie importante des parents ayant revendiqué le maintien de notes « traditionnelles » pour apprécier les progressions de la scolarité de leurs enfants, la notation a été réintroduite dans le secondaire dès la fin 2006.

En Suisse, le dossier d'évaluation mis en place dans le canton de Vaud à la fin des années 1990 a dû réintégrer la possibilité d'évaluations chiffrées en 2004, sous la pression là encore des familles. Ces motivations conduisent à une évaluation qui crée une différenciation entre élèves, là où elle était censée servir les apprentissages et faire réussir les élèves, en les considérant comme perfectibles.

E. Mangez a constaté, à l'occasion de la réforme pédagogique en Communauté française de Belgique, une même sensibilité extrême à la question de la notation, donnant lieu à des modalités de mise en œuvre très différentes selon la place des établissements dans la hiérarchie sociale et scolaire (Mangez, 2008). En bas de la hiérarchie, les codes couleur de type « feux de circulation » (rouge, orange, vert) sont utilisés aussi bien pour l'évaluation du comportement vis-à-vis du travail scolaire (participation, bonne volonté, sérieux, efforts...) que pour l'évaluation des aspects cognitifs de l'apprentissage (savoirs acquis). En haut de la hiérarchie, en revanche, la pression est forte (en particulier provenant des parents) pour le maintien d'une évaluation chiffrée et il est observé un certain refus, y compris par les élèves, des évaluations « qui ne comptent pas » (la question de l'évaluation formative étant réduite au débat sur le fait de noter ou de ne pas noter).

Il faut relever que la centralité de la note n'a rien d'inéluctable. Dans plusieurs pays, les systèmes de notation sont différents et les notes n'apparaissent que relativement tard dans la scolarité. Ainsi, tout près de la France, le Danemark ne connaît pas la pratique de la notation jusqu'à l'équivalent de la fin de notre collège. En Suède, la notation, sous forme d'échelle de A à F (E étant la note de passage) ne commence qu'à partir de la 6<sup>e</sup> année scolaire (après avoir longtemps débuté en 8<sup>e</sup> année (équivalente à la 4<sup>e</sup> française).

En France, on oublie souvent que la notation chiffrée n'apparaît, dans les pratiques des enseignants, qu'au cycle 3 (deux dernières années du primaire), afin de préparer les élèves au collège et dans un souci de cohérence avec le secondaire. Cette notation intervient alors dans les enseignements fondamentaux (orthographe, conjugaison, calcul mental) et en histoire et sciences. « *On peut donc dire que la notation chiffrée apparaît, à l'école primaire, comme un épiphénomène dans ce qui différencie le système français par rapport à d'autres systèmes éducatifs* » (Houchot *et al.*, 2013). L'alternative à la notation prend la forme d'appréciations littérales, inscrites dans le livret scolaire ou le livret de compétences (LPC). Elles permettent de faire « *apparaître les progrès des élèves, de les guider dans les régulations nécessaires sans stigmatiser les échecs* », même si parfois peu explicites, voire « *plus dévalorisantes que la note chiffrée* » pour les parents et les élèves (Houchot *et al.*, 2013). Par contre, si le codage couleur est le plus souvent utilisé pour les petites classes pour faire place à un code plus précis pour le cycle 3, il existe des écoles où aucune homogénéité n'a pu être trouvée entre enseignants non seulement d'un même cycle mais pour un même niveau. Pas de réflexion collective, donc, et pas d'intervention des directeurs pour cause de liberté pédagogique.

### La note : un outil peu précis mais blessant

On voit donc que la question de la notation reste sensible, malgré la profusion d'études qui soulignent à l'envi son caractère précaire et peu objectif, lié à la subjectivité

inévitable du jugement de l'enseignant (Suchaut, 2008).

En effet, le jugement de l'enseignant n'est pas la simple traduction des performances effectives des élèves, comme le montrent de nombreuses études dans le domaine de la psychologie sociale (Bressoux & Pansu, 2003). D'autres éléments entrent en jeu. C'est le cas de certaines caractéristiques comme l'origine sociale, le genre et le retard scolaire de l'élève. Quelle que soit la « bonne » volonté de l'enseignant, il n'évalue pas de la même façon l'enfant d'employé ou celui de cadre, ni la jeune fille ou le jeune garçon, ni le redoublant ou le « bon » élève à l'heure. L'évaluation anonyme pure étant finalement rare, tous les éléments de connaissance directe ou indirecte de l'élève participent à former le jugement de l'enseignant (et pas forcément de façon négative, d'ailleurs).

P. Merle, qui a mené une enquête minutieuse sur la « fabrication » des notes dans l'enseignement secondaire (2007) souligne aussi l'adhésion professorale à l'idée de stabilité des compétences scolaires d'une année à l'autre : le mauvais élève est voué à le rester, le brillant à confirmer sa réussite, celui dont le grand frère était « médiocre » va probablement marcher dans ses pas, etc. Selon le chercheur, on retrouve à l'œuvre ces « stéréotypes évaluatifs » dans toutes les expériences de contrôle des corrections. Un tel jugement est également influencé par le contexte de la classe : plus cette dernière est « forte » par exemple, plus le jugement de l'enseignant est globalement sévère. Or, ces constatations ne constituent pas des accidents mais bel et bien des phénomènes récurrents et réguliers, au-delà des contextes nationaux (Lafortune & Allal, 2008).

Par ailleurs, les notes sont encore le fruit d'un « bricolage » ou d'un arrangement qui est inhérent à leur processus même de fabrication, souligne P. Merle, qui montre comment de nombreuses situations d'évaluation sont « éclairées » par l'examen de la scolarité antérieure, du livret scolaire et de la prise en compte de la réputation des établissements déjà fréquentés.

L'analogie souvent faite avec le « thermomètre » apparaît donc bien peu pertinente, sauf à considérer que l'important est moins



la température qu'on prétend mesurer que le fait d'en attribuer une quelconque à chacun.

Pour Merle (2012a), on peut néanmoins tracer des pistes d'amélioration de la notation :

- recours aux barèmes pour limiter les aléas ;
- préférer les épreuves communes souvent plus équitables ;
- préserver l'anonymat scolaire et social de l'élève ;
- favoriser un usage « encourageant » de la note ;
- contractualiser et expliciter les pratiques d'évaluation ;
- distinguer évaluation formative et sommative ou certificative ;
- améliorer la cohérence entre socle commun, programmes, enseignements et évaluation ;
- s'inspirer d'échelles de notation d'autres pays qui s'avèrent moins décourageantes pour les élèves.

Pour de nombreux acteurs sociaux, les classements sont un mal nécessaire dans une société inégalitaire. Dès lors, l'enjeu des évaluations scolaires est de se rapprocher autant que possible de principes d'équité et de justice pour que les classements scolaires, qui donnent accès à des positions sociales différenciées, soient aussi clairs et explicites que possible aux yeux des parents, les élèves et les enseignants (Maulini, 2012).

### La constante macabre

Comme le souligne Antibi (2003), la notation, telle qu'elle est appliquée actuellement, est régie par des principes d'usages tels que les performances des élèves sont plus basées sur les échecs que sur les réussites. Cela vient d'une croyance durable en l'impossibilité de n'avoir que des bons élèves et donc en la suspicion marquée envers une moyenne de la classe trop élevée. Il est de bon ton de donner des mauvaises notes. L'enseignant, poussé par le corps d'inspection, les collègues ou ses propres croyances, va faire en sorte d'établir des critères de notation lui permettant d'obtenir une moyenne entre 8 et 12. Il va donc construire les exercices et devoirs évalués dans ce sens, plutôt que d'opter pour une stratégie visant à motiver les élèves et évaluer les acquisitions plutôt que les manques.

### L'exemple de la dictée

La dictée représente un exercice majeur pour la maîtrise de l'orthographe. Ce n'est pas le « seul exercice d'appropriation de l'orthographe », mais elle offre l'avantage d'une concentration spécifique sur l'orthographe et ses règles, là où d'autres productions écrites doivent laisser le champ libre à l'expression et l'inventivité. « *Un élève ayant à gérer simultanément la réflexion, la recherche de l'expression, la graphie et la mise en page de son écrit, et le souci de la correction de la langue, se trouve ainsi confronté à des tâches mentales multiples, dont on sait qu'elles ne peuvent solliciter également son attention* » (Barbarant, 2014).

Contrairement à d'autres domaines d'enseignement pour lesquels la démarche sommative consiste donc à additionner des éléments (représentant des niveaux de performance), l'enseignement de la langue utilise une démarche sommative soustractive, le correcteur opérant une soustraction pour chaque incorrection relevée (la fameuse « faute d'orthographe »). Lors de l'évaluation de production de textes, l'enseignant va à la fois opérer une évaluation sommative additive (sur le fond) et soustractive (maîtrise linguistique). La dictée offre donc l'avantage de n'avoir qu'un type de connaissances à évaluer. Pour autant, l'objectif en cours d'apprentissage n'est pas de sanctionner (valider un état de fait), mais identifier, décrire, améliorer les apprentissages. Pour l'élève en difficulté, cette sanction qui, pour le coup, révèle le côté pénalisant du mot, sera immanquablement disqualifiante et le sentiment de nullité ressenti par l'élève risque fort de le démotiver plutôt que l'inciter à s'améliorer.

En 2013, la correction de la dictée du diplôme national du brevet a été l'occasion d'expérimenter un barème graduel de notation. Si les erreurs sont toujours comptabilisées et mises en évidence par le correcteur, les points acquis sont basés sur le décompte des graphies correctes. L'enjeu affiché est de distinguer les compétences orthographiques des élèves, « *de hiérarchiser les difficultés, d'envisager des remédiations*

personnalisées, adaptées aux capacités distinctes des élèves » ([Éduscol](#)). On passe d'une évaluation « descendante » à une évaluation « positive ».

Le projet de nouveau barème s'est inscrit dans une démarche visant à traiter de la difficulté scolaire, à faire évoluer l'évaluation et à s'insérer ultérieurement dans le cycle CM1-CM2-6<sup>e</sup>. Il a été expérimenté dans l'académie de Poitiers, en juin 2013, où 1 563 copies ont pu être analysées.

Concrètement, les erreurs orthographiques sont catégorisées en « syntaxe du groupe verbal », « syntaxe du groupe nominal » et « orthographe lexicale ». Chaque enseignant a à disposition un diaporama et un tableur contenant des consignes pour chacune des catégories d'erreurs ou plus exactement des réussites et des erreurs. Il a en main des éléments portant sur les difficultés proposées par le texte et peut envisager la progression pédagogique nécessaire.

L'objectif final de ce tableur est bien de noter l'épreuve du brevet et l'application s'appuie sur un ensemble de critères pondérés. Au quotidien, cela peut permettre de prendre en considération des difficultés diverses, de définir des niveaux de valeurs orthographiques ou grammaticales, même si certains enseignants ont trouvé le système lourd et chronophage. Pour des élèves ayant obtenu un zéro, avec la notation classique, la correction par barème graduel a permis d'identifier des profils de compétences différents.

Cette méthode n'empêche en aucune manière de conserver une correction portant sur les « fautes » d'orthographe, dans le cadre d'une évaluation formative, au cours de laquelle hiérarchisation et repérage des difficultés, exercices supplémentaires permettent de poser un nouveau diagnostic (Gerard, 2013 ; De Ketele, 2013).

Comment évalue-t-on l'orthographe dans les autres pays ? Il est difficile de trouver des travaux de recherche portant sur la dictée (*dictation*) et sur l'orthographe et son évaluation. En France les compétences orthographiques sont envisagées dans le contexte de production d'écrits : l'attention portée à l'orthographe peut

brider l'imagination et la créativité narrative de l'élève ; inversement les élèves ne prêtent que peu d'attention à l'orthographe et la grammaire lorsqu'ils rédigent des textes. Ailleurs, les travaux portent sur les liens entre énonciation (*spelling*), écriture (*writing*) et lecture (*reading*) et il est rare de trouver une dissociation entre ces trois « compétences ».

## LES ÉVALUATIONS STANDARDISÉES COMME RITUELS SANCTIONNANT LA FIN D'UN CYCLE

L'évaluation certificative représente une grande partie des évaluations diplômantes. Elle a pour fonction de certifier la maîtrise des apprentissages, en général pour vérifier que les objectifs assignés à un cycle ont été atteints et que les compétences visées sont acquises.

On confond souvent l'évaluation certificative et l'évaluation sommative fondée sur une mesure globale attestée par une note lors d'un examen externe, alors même qu'on peut imaginer qu'une évaluation certificative s'appuie sur des appréciations progressives en cours d'apprentissage (De Ketele, 2013 ; Gérard, 2013). Ainsi, le contrôle continu en cours de formation dans la voie professionnelle est un exemple, d'ailleurs controversé, d'évaluation certificative qui s'appuie sur une grande part de contrôle continu.

Ceci noté, la plus grande part des évaluations certificatives ressemble à des examens externes, même s'il est difficile de trouver un examen de passage ou validant une fin de scolarité commun dont la recherche ait pu faire état et dont les modalités de mise en œuvre puissent être comparées.

On sait en fait très peu de choses sur les examens et il est d'ailleurs significatif qu'on ne dispose que de très peu d'études actuelles sur le baccalauréat ou le brevet, provenant de la recherche (centrée sur les pratiques évaluatives des enseignants) ou de la DEPP (qui préfère se concentrer sur ses propres évaluations diagnostiques).

Le baccalauréat est une énorme machine, qui mobilise des moyens humains et fi-



nanciers énormes, qui ampute les cours d'au moins un mois, mais dont on ne sait finalement pas grand chose d'un point de vue scientifique. Quelques exemples permettent néanmoins de mettre en balance le baccalauréat français avec d'autres instruments certificatifs de fin d'enseignement secondaire.

### Le baccalauréat, monument en perte de sens

Comme chaque année, à l'approche des épreuves du baccalauréat, les plus concernés comme les moins informés se déchainent quant à l'utilité et l'efficacité du baccalauréat. Les docimologues ont montré, depuis les années 1930, le caractère aléatoire de la notation des épreuves du baccalauréat (Suchaut, 2008). L'application d'une grille de correction plus précise, par exemple, ne suffirait pas à réduire l'incertitude de la notation et à rendre équitable l'obtention ou non du diplôme. La question tient donc dans la pertinence actuelle du baccalauréat. Les valeurs qui ont longtemps légitimé cet examen, sélectivité, accès aux études supérieures, accès à l'emploi (public) sont désormais obsolètes. Le bac est encore en théorie le premier diplôme de l'enseignement supérieur mais, dans les faits, la procédure de [l'admission post-bac](#) officialise la généralisation progressive du recrutement par dossier. Les professeurs d'université, présidents des jurys, n'interviennent plus dans cet examen devenu examen de fin de secondaire.

Par ailleurs, il n'existe plus un baccalauréat national unique et indivisible mais bien au moins trois baccalauréats, dont l'objectif était, pour le ministre de l'Éducation nationale (Chevènement, en 1985) ayant mis en place ces trois filières « *d'offrir à l'issue de la classe de troisième, trois voies "d'égale dignité"* ».

Ces dignes voies se sont transformées en parcours de préparation du baccalauréat plutôt qu'en lieux d'éducation et d'apprentissage. Une filière, la voie S de la série générale, concentre les meilleurs élèves et fait figure de voie d'accès aux parcours les plus convoités dans l'enseignement supérieur, sans grand rapport avec la nature supposée scientifique des enseignements.

Plus encore, chacun s'interroge sur le fait qu'« *avec le baccalauréat, l'exigence évaluative est devenue prosternation devant un monument national* » (Hadji, 2012). Pour certains, le baccalauréat « *est censé porter toutes les vertus humanistes et égalitaires. Il est assimilé à la protection de tous contre l'arbitraire. [...] En Italie, la "Maturità" est, elle aussi, très contestée. Et en Allemagne, par exemple, l'Abitur comprend 70 % de contrôle continu... On ne regarde pas ce qui se passe à l'étranger, sinon on se rendrait compte que le bac n'est pas universel* » (Fize, [2013](#)).

Quels sont les arguments avancés pour défendre le baccalauréat ?

- « le bac est la sanction de la fin de scolarité » : assertion démontée par le bachotage, par les pratiques des autres pays ;
- « le bac est un rite de passage » : non, car le rite de passage est une série d'épreuves qui fait passer toute une génération de l'adolescence à l'âge adulte, mais à partir d'épreuves collectives et identiques pour tous ; les discours de démocratisation de l'école ont fait long feu face à la multiplicité de baccalauréats (par exemple) ;
- « on est fier d'avoir son bac » : on déchanté rapidement puisque ce diplôme ne donne accès qu'à des petits statuts, voire au chômage.

Il existerait d'autres motifs « supposés » de la persistance du baccalauréat à la française :

- la défense et le besoin de valorisation de certaines disciplines scolaires ;
- la culture de la méritocratie et de la sélection (il reste 15 % des jeunes qui n'obtiennent pas le bac, et certains ont un meilleur bac que d'autre) ;
- et peut-être aussi, tout simplement, la capacité d'inertie du système et de ses acteurs ?

On s'interrogera encore quelques temps sans doute sur ce hiatus entre l'existence d'un consensus pour dénoncer le coût et l'inutilité (ou au moins l'inefficacité) du baccalauréat et la persistance de cette spécificité très française.



## Des évaluations certificatives, ailleurs

La plupart des pays européens valident les apprentissages acquis en fin de scolarité obligatoire par un certificat de fin de scolarité. Ce certificat est délivré sur la base des résultats du contrôle continu, souvent sur les trois dernières années et/ou à la suite de tests portant sur les disciplines obligatoires ou choisies. L'entrée à l'université peut être conditionnée par l'obtention d'un diplôme spécifique, attestant à la fois de l'acquisition de connaissances et compétences par matières ou domaines (sciences, humanités, arts, etc.). Cela implique le plus souvent, qu'on ne pratique pas de notation globale et de compensation entre notes et, pour ceux qui veulent continuer des études dans l'enseignement tertiaire, le risque d'être sélectionné selon leurs meilleurs résultats ou plus exactement sur les compétences et connaissances acquises (seuil variable selon les pays).

En **communauté française de Belgique**, le CESS (Certificat d'enseignement secondaire supérieur) est une épreuve certificative externe commune. L'élaboration de chaque épreuve, des consignes de passation, de correction ainsi que les critères de réussite sont définis par un groupe de travail, composé de membres du Service général du pilotage du système éducatif, de l'inspection, de conseillers pédagogiques et d'enseignants. Ces épreuves (en histoire et en français) viennent compléter celles passées dans le cadre du certificat d'enseignement (CE1D) délivré « *au terme de la troisième étape du continuum pédagogique* ». Ces dernières concernent les mathématiques, le français, les sciences, les langues modernes. Pour les élèves n'ayant pas satisfait aux critères, le conseil de classe peut valider l'acquisition des compétences et savoirs sur la base du dossier personnel de l'élève (parcours de l'élève).

L'évaluation des élèves au **Danemark** ne peut manquer d'étonner l'observateur français puisque ce pays, peu suspect de nonchalance dans la compétition mondiale, ne connaît pas la pratique des notes avant le grade 8 qui correspond à l'âge de 15 ans ! Suite aux résultats des évaluations PISA en 2000, l'objectif a été d'améliorer le taux de poursuite d'études dans

l'enseignement supérieur, afin que 95 % des élèves poursuivent leur formation au-delà de l'enseignement secondaire.

Le *Gymnasium* est l'équivalent de notre lycée au niveau de l'enseignement secondaire supérieur. Ces trois années sont organisées dans la voie générale, commerciale ou technique, suivies par 65 % d'une génération. Les élèves reçoivent trois notes par sujet en moyenne, ainsi qu'une note finale quand la matière est terminée (notation sur une échelle de 7 points).

Au moment de la sortie du lycée, une évaluation externe et centralisée intervient dans 7 domaines : 5 obligatoires (danois, mathématiques, anglais, chimie et sciences) plus 2 domaines choisis chaque année de façon aléatoire par le ministère la veille de l'épreuve nationale (autrement dit, pas d'impasse possible). Comme certaines matières font également l'objet d'un examen à l'oral (anglais, chimie, etc.), chaque élève passe au total 9 examens.

On pourrait dire que les points de similitude avec notre baccalauréat s'arrêtent ici. Les examens danois sont en effet considérés comme des épreuves de compétences plus que de mémorisation ou de répétition. Elles privilégient par exemple des travaux sur des textes que les élèves n'ont jamais eu l'occasion d'étudier durant la scolarité ou sur des périodes historiques qui ne figurent pas au programme. Sauf rares exceptions (certaines formules de mathématiques par exemple), la structure des examens fait qu'il n'y a pas de « bonnes » réponses toutes faites à recopier.

Ces examens terminaux peuvent prendre des formes extrêmement diversifiées :

- sujet pluridisciplinaire avec interrogation orale par différents enseignants ;
- examen écrit de 4 à 5 heures avec accès à Internet ;
- 15 à 20 pages de dissertation écrite sur un projet spécialisé, à faire en deux semaines ;
- examen oral en groupe, avec notation individuelle ;
- examen oral individuel avec 30 minutes de préparation... ou 24 heures ! ;
- examen oral de présentation de projet, avec questions aléatoires de type « grand oral » sur d'autres sujets.



De façon générale, le système scolaire répercute en grande partie le principe de « confiance sociale » ●, qui se traduit en l'occurrence par un souci d'impliquer les élèves dans leurs études en créant un climat de confiance.

Aux **États-Unis**, jusqu'à la fin de l'enseignement secondaire, l'éducation générale, la socialisation et l'accomplissement personnel des jeunes sont les objectifs prioritaires (Colardyn, 2009). Ainsi, notre équivalent du lycée (*High School*), organisé autour de *credits* laissés au choix des élèves, débouche sur un diplôme peu sélectif de fin des études secondaires, qui clôt le parcours des jeunes plus qu'il n'atteste d'une qualification ou d'une maîtrise de connaissances et de compétences. La formation professionnelle se déroule plutôt après, soit dans les différents *community colleges*, soit dans les universités, après des procédures d'admission sur dossier et réussite au test d'accès à l'enseignement supérieur (*Scholastic assessment test* ou SAT). Cette place relative des diplômes de fin de scolarité explique en partie le développement d'évaluations standardisées au niveau des états et au niveau fédéral, pour assurer une certaine comparabilité des résultats des écoles.

Dans ce contexte, la liberté de choix de l'école et la mise en concurrence des établissements favorisent toute initiative permettant de valoriser son établissement. Ainsi, de plus en plus de lycées adoptent le baccalauréat international (BI). Créé par une fondation privée, ce BI propose des programmes à visée « internationale », dès l'âge de 9 ans, aboutissant à un diplôme pour les jeunes de 16 à 19 ans. Dispositif coûteux pour les établissements (de l'ordre de 10 000 dollars), le BI a été mis en avant en 2007 par une loi encourageant l'accès des plus défavorisés à l'*Advance placement* (AP) mais aussi au BI. Resnik montre que le BI est avant tout un programme donnant accès aux meilleures universités et qui bénéficient à une minorité de très bons élèves (Resnik, 2013).

Pour décrire la situation en **Finlande**, il convient de rappeler que l'enseigne-

ment de base, qui comprend le primaire et le secondaire, est sans rupture de parcours comme en France. Il n'y a pas d'examens nationaux. L'évaluation finale, sur la base de laquelle les élèves seront sélectionnés pour poursuivre d'autres études doit être comparable à l'échelle nationale et traiter équitablement les élèves. En fin du cycle secondaire supérieur, les élèves passent un examen de fin d'études (*ylioppilastutkinto*) comportant quatre épreuves dont l'une, de langue maternelle du candidat, est obligatoire pour tous. Les trois autres épreuves sont au choix : deuxième langue nationale, langue étrangère, mathématiques, sciences humaines (culture générale). Le candidat peut passer d'autres épreuves facultatives. L'objectif de cet examen est de vérifier que les élèves ont acquis les connaissances et compétences du programme d'études secondaires supérieures et atteint la maturité et la légitimité suffisante pour une poursuite d'études en conformité avec les objectifs. Pour mémoire, la fin de scolarité obligatoire (un peu plus tôt) est balisée par un certificat attestant de l'atteinte des objectifs fixés par le programme de l'enseignement de base, remis par l'école (évaluation concertée de l'ensemble des enseignants).

En **Allemagne**, la fin de la scolarité gymnasiale (lycée général) est sanctionnée par un examen, l'*Abitur*, dont la responsabilité incombe au ministère de l'Éducation de chaque *Land* ; il est de plus en plus rare que les lycées soient à l'origine des épreuves, mais ils restent en charge de l'organisation et des corrections de leurs élèves. Les notes obtenues à l'*Abitur*, associées à celles du contrôle continu des deux dernières années du lycée, servent à établir le certificat d'entrée dans l'enseignement supérieur – *Allgemeine Hochschulreife*. Si le contrôle continu est jugé insatisfaisant, l'élève peut redoubler une fois (Endrizzi, 2013).

En **Espagne**, l'éducation secondaire obligatoire (ESO) comprend quatre années de 12 à 16 ans. Un diplôme de l'enseignement secondaire est délivré par un conseil d'enseignants selon les résultats obtenus par les élèves lors de l'ESO (évaluation sous forme

Le concept de « confiance sociale » (*social trust*) a été développé par le politiste danois Gert Tinggaard Svendsen, chercheur à l'université d'Aarhus (DK).

de contrôle continu). En cas d'échec, une certification officielle constatant les compétences acquises peut être délivrée. Le diplôme leur permet d'accéder au cycle préparant au *bachillerato* (choisi par plus de 70 % des élèves) ou à la formation professionnelle appelée *ciclo formativo de grado medio*. Le *bachillerato* est accordé aux élèves qui ont une évaluation positive dans toutes les disciplines, durant les deux années (soit 5/10). Cette décision est prononcée par un conseil d'enseignants à la fin de la deuxième année. Les 3 filières du *bachillerato* (arts, sciences et technologie, humanités et sciences sociales) ouvrent la voie de l'enseignement supérieur, mais les futurs étudiants doivent également réussir la *Prueba de Acceso a la Universidad* (aussi appelée *selectividad*), dont les principes généraux sont donnés par un décret royal et dont l'organisation relève d'une commission composée entre autres d'universitaires et d'enseignants de *bachillerato* et nommée par les Communautés autonomes (Endrizzi, 2013).

En **Suède**, la fin de la scolarité obligatoire (16 ans) est marquée par l'obtention d'un certificat de fin d'études secondaires basé sur le contrôle continu des trois dernières années. En 2011, plus de 82 % des 25-64 ans possédaient ce certificat, donnant à la Suède le plus haut niveau d'éducation secondaire des pays de l'OCDE.

## ÉVALUER POUR FORMER

La littérature de recherche francophone est essentiellement concernée par l'évaluation DES apprentissages. La question est de savoir si on doit évaluer pour rendre compte de ce qui a été appris (évaluation DE l'apprentissage) ou évaluer pour prendre des décisions affectant, à court terme, l'enseignement et l'apprentissage (évaluation POUR les apprentissages), voire même évaluer pour étayer l'apprentissage, par exemple en apportant à celui qui apprend une meilleure maîtrise du processus d'apprentissage.

« L'évaluation pour les apprentissages permettra d'ajuster l'enseignement pour faire progresser l'apprentissage dans le sens attendu, à condition toutefois que les outils d'évaluation soient suffisamment bien construits pour apporter de l'information dans ce sens » (Rémond, 2008).

S'interroger sur l'évaluation pour les apprentissages, c'est se situer à rebours d'une tradition où l'évaluation est conçue dans un modèle cybernétique comme un contrôle à posteriori, quasiment technique, d'un apprentissage qui serait indépendant des modalités d'évaluation. Un grand nombre de modèles d'évaluation se situent dans cette perspective cybernétique de façon explicite ou implicite (Vial, 2012).

Explorer l'évaluation pour la formation, c'est au contraire essayer de dépasser l'évaluation comme un instrument de contrôle de conformité à une sorte de parcours de réussite pédagogique modèle, défini ailleurs, pour un élève « moyen » (moyen étant entendu comme la représentation qu'on se fait à un moment donné de ce que doit être un élève moyen).

L'évaluation formative est ainsi vue comme une part du processus d'enseignement et d'apprentissage, plutôt que comme une activité séparée, intervenant après une phase d'enseignement (Looney, 2011). Elle cherche avant tout à favoriser l'appétence et la motivation des élèves, en leur apportant un retour d'informations constant sur leur apprentissage mais dans un contexte favorisant leur sentiment d'efficacité personnelle, pour reprendre la théorie de Bandura ●.

On consultera à ce sujet avec profit les articles d'Alain Lieury parus sur le site des Cahiers pédagogiques en [juillet 2014](#).



## ÉVALUER ET FORMER

### Les principes de l'évaluation formative

Souvent mobilisés pour aborder la notion d'évaluation formative ou d'évaluation formatrice, les travaux de Black et William (1998) proposent dix principes pour

guider la pratique évaluative pour les apprentissages, en classe. En croisant leurs premières analyses, à la fin des années 1990, avec des textes plus récents, nous aborderons les processus d'évaluation pour les apprentissages, les éléments constitutifs et les prolongements en matière de pratiques enseignantes.

Les principes de l'évaluation pour l'apprentissage (Black & William, 1998)	Les principes de l'évaluation formative (CERI/OCDE)
<ul style="list-style-type: none"><li>- elle est constitutive d'une planification efficace pour l'enseignant</li><li>- elle se centre sur la manière dont les élèves apprennent</li><li>- elle est centrale à la pratique pédagogique</li><li>- c'est une compétence professionnelle essentielle de l'enseignant</li><li>- elle est progressive et constructive</li><li>- elle renforce la motivation des élèves</li><li>- elle sert la compréhension des objectifs pédagogiques et des critères d'évaluation par les élèves</li><li>- elle aide les élèves à améliorer leurs compétences</li><li>- elle développe la capacité chez l'élève à s'auto-évaluer</li><li>- elle reconnaît toutes les formes de réussite scolaire</li></ul>	<ul style="list-style-type: none"><li>- instauration d'une culture de classe qui encourage l'interaction et l'utilisation d'outils d'évaluation</li><li>- définition d'objectifs d'apprentissage et suivi des progrès individuels des élèves vers ces objectifs</li><li>- utilisation de méthodes d'enseignement variées pour répondre aux besoins diversifiés des élèves</li><li>- recours à des méthodes diversifiées pour évaluer les acquis des élèves</li><li>- feedback sur les performances de l'élève et adaptation de l'enseignement pour répondre aux besoins identifiés</li><li>- implication active des élèves dans le processus d'apprentissage</li></ul>

L'évaluation formative contribue au processus continu d'apprentissage. Cette assertion éclairante quant à la nécessité de se centrer sur les apprentissages de l'élève et non sur des contraintes externes, quelles qu'elles soient, ne suffit pas à guider l'enseignant dans sa pratique pédagogique.

Les travaux de recherches récents décrivent la diversité des méthodes d'évaluation formative et évoquent les notions de régulation, de rétroaction (*feedback*). Il est par exemple important d'apporter à l'élève un *feedback* circonstancié et approprié aux moyens de franchir une étape ou de surmonter un obstacle (« comment faire »). Au contraire, un retour général centré sur la personne ou

l'effort mais déconnecté de l'objectif cognitif peut avoir un impact négatif à terme. Il en est ainsi des retours formulés en termes de « il faut plus travailler », « le problème n'a pas été compris », « tu as bien travaillé », etc. (Looney, 2011).

Allal et Mottier Lopez (2005) ont étudié l'évolution des concepts entre ceux décrits par Bloom (sur la pédagogie de maîtrise, en 1978) et ceux utilisés dans la littérature de recherche francophone entre 1978 et 2002. Dans l'approche élargie (selon les termes d'Allal et Mottier Lopez), l'évaluation formative se situe à n'importe quel moment de l'apprentissage et non pas à la suite d'une phase d'enseignement ; elle se fait à l'aide

de moyens d'information variés et non plus sous forme de tests ; le *feedback* qui s'en suit peut induire une adaptation de l'enseignement et ne donne pas uniquement lieu à une correction du travail de l'élève ; les élèves participent à ces étapes évaluatives, alors que chez Bloom, l'évaluation formative est gérée par l'enseignant. Là où Bloom recherche la maîtrise des objectifs par tous les élèves, la recherche récente met en avant la différenciation de l'enseignement, voire des objectifs. Enfin, la remédiation, en cours d'apprentissage ou après évaluation, est remplacée par la notion de régulation, pour les élèves et pour l'amélioration des pratiques pédagogiques. Allal et Mottier Lopez distinguent trois types de régulation, interactive, rétroactive (remédiation) et proactive.

Des approches postérieures à cette revue de littérature mêlent ces trois types de régulation, les adaptant aux différentes phases de l'enseignement/apprentissage et en combinant les dispositifs évaluatifs (diagnostique/formatif/sommatif). La rétroaction peut provenir de l'enseignant ou d'un autre élève, elle peut être orale et se faire en classe entière, ou dans des petits groupes ; elle peut être écrite, selon différents types de support. L'essentiel est de donner des informations qui permettent à l'élève de modifier ou d'ajuster son travail, sur la base de critères explicites et décrits préalablement. Les commentaires, éléments essentiels de cette rétroaction se doivent d'être positifs, constructifs, spécifiques, liés aux résultats d'apprentissage ; amener des suggestions, elles aussi constructives et positives ; être faits opportunément (Allal & Mottier Lopez, 2005 ; Hattie & Timberley, 2007).

M. Crahay (2007) a réalisé une revue de littérature portant sur le processus de *feedback*, depuis la fin des années 1960. Il effectue en quelques tableaux et commentaires des typologies des *feedbacks* et leurs conséquences sur les apprentissages et le comportement des élèves. De Landsheere ou Bayer décrivent les postures évaluatives positives et négatives et l'intervention de l'affectif dans ces postures. On peut citer par exemple une évaluation qui consistera à approuver en répétant la réponse de l'élève (positive) ou à la répéter de façon ironique (négative) ; l'une encourage là où l'autre est critique. Cette évaluation simple

demande ensuite un développement (réaction) : faire compléter la réponse, la corriger s'il y a lieu, clarifier, expliciter, synthétiser.

Dans ces schémas reprenant les réactions des enseignants pour expliciter une évaluation, on observe différentes attitudes, allant d'une réaction sans explication (forme de verdict), demandant une correction sans indication et renvoyant l'élève à une cause non contrôlable, à une réaction avec explication, donnant des indices pour la correction et renvoyant l'élève à une cause contrôlable. Le fait de compléter la communication simple (approbation ou désapprobation) en demandant un développement à l'élève (correction ou amélioration de la réponse) permet de réguler l'apprentissage, de même qu'un développement de la part de l'enseignant, lors d'une réponse partiellement correcte ou incorrecte.

Le propos de Crahay n'est pas de décliner les résultats de recherche mais d'émettre des réserves quant à la scientificité des conclusions émises. La pertinence des réactions observées ne les rend pas généralisables pour autant. Le contexte peut induire d'autres effets bénéfiques ou non : situations d'apprentissage, type de tâches, personnalité de l'élève, progressivité des apprentissages.

Si l'on s'intéresse à ce dernier élément, à savoir la manière dont les enseignants organisent la progression des apprentissages, une autre donnée vient infléchir la réaction des enseignants : l'origine socio-économique des élèves et leurs attitudes (participatifs pour les mieux nantis et plus craintifs pour les autres). L'enseignant complétera l'évaluation par des situations-problèmes dans le premier cas ou bien « *se cantonnera dans une présentation ex cathedra de la manière permettant un contrôle serré de la progression des élèves* » (Crahay, 2007). On peut alors s'interroger sur le bien-fondé d'une moindre ambition de la part des enseignants envers les plus modestes et sur la réalité d'une réassurance des élèves les plus anxieux.

Crahay préconise « *d'investiguer les réactions par lesquelles les enseignants peuvent favoriser chez les élèves la mise en place aussi régulière que possible de procédures d'autorégulation, tout en engendrant des effets substantiels d'appren-*



tissage disciplinaire », quels que soient l'origine socioculturelle, le profil motivationnel des élèves.

Diverses études ont aussi montré comment des enseignants peuvent être amenés à multiplier les concessions lors des évaluations sommatives en classe pour ne pas décourager ou démotiver les élèves en difficulté scolaire (Millon-Fauré, 2013). L'enjeu essentiel est alors de préserver la relation didactique dans la classe, au prix d'un aménagement des règles théoriques de l'évaluation sommative et d'une confusion de fait entre dimension « formative » et dimension « sommative ».

### Comment agencer les types d'évaluation dans les temps d'apprentissage

« L'utilisation des termes "évaluation formative" et "évaluation sommative" pouvaient donner l'impression qu'il y avait diverses formes d'évaluation ou encore qu'elles étaient liées à des méthodes variées de recueillir l'information. Ce n'est pas le cas ; ce qui importe, c'est l'intention de l'évaluation » (Harlen, 2006, cité par le ministère de l'Éducation de l'Ontario, 2010).

Les différents types d'évaluation pourraient s'articuler de la manière suivante, selon un schéma dicté par l'intention d'évaluation et avec pour objectif l'exploitation de l'information recueillie lors de l'évaluation, en accord avec l'intention (Stobart, 2011). Ainsi, si l'on considère comme intention préalable une évaluation au service de l'apprentissage ●, l'évaluation diagnostique, utilisée en début de période d'enseignement-apprentissage, sera complétée par une évaluation formative, pratiquée de manière continue pendant l'acquisition des connaissances et habiletés. L'étape diagnostique va permettre de collecter des informations sur les connaissances antérieures de l'élève, ses attentes quant aux connaissances et habiletés à acquérir, de planifier une différenciation appropriée (par rapport aux objectifs) de l'enseignement, de l'apprentissage et de l'évaluation. La deuxième étape va permettre de suivre les progrès de l'élève, de communiquer et décrire cette progression à l'élève et d'envisager avec lui les stratégies à adopter.

Cette phase de rétroaction favorise l'auto-régulation chez l'élève, et l'amène à comprendre où il en est dans le processus

d'apprentissage. L'intention évaluative est ici d'utiliser l'évaluation en tant qu'apprentissage. « C'est un moyen efficace de gérer l'utilisation et le développement de stratégies nécessaires à l'atteinte des objectifs fixés » (Ministère de l'Éducation de l'Ontario, 2009). Là encore, l'évaluation formative, fréquente, va évaluer les stratégies utilisées par l'élève, sa capacité à réfléchir à son apprentissage et aux moyens qu'il se donne pour s'améliorer. Construite à partir de critères préétablis (collectivement ou individuellement), cette phase peut se faire en autoévaluation ou en évaluation par les pairs ●.

En troisième intention, on passera à une évaluation de l'apprentissage, qui sert à confirmer ce que l'élève connaît et sait faire. Cette étape sommative sera accompagnée de commentaires « précis et opportuns » de l'enseignant. Elle se fera à la fin de chaque unité d'apprentissage ou lorsque l'élève aura montré qu'il a acquis suffisamment de connaissances ou compétences. Pour être pertinente (ou efficace), elle doit être accompagnée d'un échange avec l'élève quant aux résultats d'apprentissage, aux modalités pour atteindre ces résultats et aux objectifs pour réutiliser ces connaissances dans les prochains cours ●.

Ce processus n'a rien d'évident, car les enseignants recourent plus spontanément à l'apprentissage par cœur et ne formulent que des questions superficielles pour vérifier l'apprentissage de leurs élèves. Ils rencontrent en outre des difficultés à dépasser des retours en termes superficiels, à interpréter les réponses de leurs élèves, de nature à construire les prochaines étapes qui seraient nécessaires pour progresser dans leur apprentissage (Herman *et al.*, 2010).

Pour certains, l'évaluation formative doit faire partie intégrante de l'acte d'enseigner, elle s'intègre alors de manière dynamique dans le déroulement d'une séquence pédagogique et doit être opportunément complétée par des phases d'auto-évaluation ou d'évaluation entre pairs. Ces phases d'évaluation par les élèves permettent de reboucler sur le point de départ de l'évaluation que sont les objectifs (communs) d'apprentissage (Oswalt, 2013).

« L'évaluation au service de l'apprentissage vise à recueillir et à interpréter les preuves d'apprentissage afin de permettre tant au personnel enseignant qu'à l'élève de déterminer l'apprentissage ciblé, d'établir où l'élève se situe dans son apprentissage, et déterminer ce qui doit être fait pour y arriver » (Assessment Reform Group, 2002, cité par le ministère de l'Éducation de l'Ontario).

On trouvera, sur le site du *National Council for Curriculum and Assessment* (Irlande), plusieurs vidéos montrant des situations d'évaluation. L'une d'entre elles montre un temps d'évaluation par les pairs, après un exercice de lecture à voix haute ([AFL toolkit - more techniques](#)).

En 2009, le ministère de l'Éducation de l'Ontario a réalisé un guide « [Évaluer, différencier... réussir !](#) » reprenant cette articulation des différents types d'évaluation et donnant des exemples de commentaires et tâches à envisager aux différentes phases d'évaluation.

Cela suppose que l'explicitation de la part de l'enseignant, sur les procédures à mettre en œuvre (stratégies de réalisation) et sur les critères de réussite, ne soit pas donnée *ex abrupto*. L'élève doit en effet s'approprier ces critères, prendre du recul, comparer ce qu'il a fait avec ce qu'il aurait du faire, autrement dit évaluer son travail par rapport à ce que Nunziati appelait les dénominateurs des critères de réussite, à savoir la pertinence, la complétude et l'exactitude et, selon les tâches, le volume des connaissances et l'originalité. Cette expertise doit se construire; or, « *c'est une compétence de haut niveau taxonomique qui met du temps à s'acquérir* ». Pour pallier cette réflexion chronophage, on peut ménager des temps d'élaboration des critères en commun. On passe ainsi d'un processus d'évaluation formative, temps d'intervention-régulation, à une phase d'évaluation formatrice où l'élève s'auto-évalue,

s'auto-régule et prend conscience de ce qu'il sait et doit faire pour réussir (Nunziati, 1990, citée par Castingaud & Zakhartchouk, 2014).

### La mesure de l'évaluation formative

Pour vérifier l'efficacité d'une évaluation formative sur les apprentissages *in situ*, il faut pouvoir observer, en classe, la présence de trois des principales composantes de l'évaluation formative : les objectifs d'apprentissage, le contrôle de ces apprentissages, et le *feedback*. Plusieurs travaux nord-américains proposent des séries de questions permettant d'évaluer l'évaluation formative, c'est-à-dire d'évaluer le processus par lequel enseignant et élèves produisent une rétroaction et ajustent leur enseignement et/ou leurs apprentissages (Oswalt, 2013 ; Airasian & Russel, 2008).

Les (bonnes) questions que l'enseignant doit se poser (Oswalt, 2013) :

- s'est-il assuré de la bonne compréhension par les élèves des objectifs d'apprentissage pour la session ? Pour la tâche à accomplir ?
- donne-t-il des exemples de qualité ?
- a-t-il abordé d'éventuels malentendus quant aux objectifs de réussite de la tâche ?
- fait-il des efforts pour suivre l'apprentissage des élèves en continu ?
- propose-t-il aux élèves une variété de possibilités et de méthodes pour répondre aux questions ?
- utilise-t-il des stratégies efficaces de questionnement (temps d'attente adéquate, questions ouvertes) pour s'assurer des apprentissages ?
- vérifie-t-il l'acquisition à la fois de connaissances factuelles/procédures et conceptuelles ?
- vérifie-t-il les capacités de transfert de ces connaissances au sein et entre les disciplines ou domaines ?
- permet-il à l'élève de réagir à la suite (ou au cours) d'une phase d'évaluation ?
- fournit-il des commentaires précis qui aident l'apprentissage ?
- utilise-t-il une base critériée d'évaluation pour éviter toute comparaison entre élèves ?
- propose-t-il des possibilités et des stratégies alternatives d'amélioration ?
- informe-t-il l'élève sur ses atouts et les stratégies pour persévérer ?

Les partisans inconditionnels de l'évaluation formative s'appuient sur les effets mesurés par des méta-analyses d'expériences ou dispositifs d'évaluation formative. Des scores de progression des élèves à des tests standardisés sont ainsi parfois mis en exergue pour attester de l'intérêt de dispositifs ayant été mis en place les préconisations de Back et Wiliam.

Si les « preuves » de nature quantitative et empirique constituées par ce type de mesures sont séduisantes, on peut douter qu'elles soient toujours appropriés pour juger de la pertinence de l'évaluation formative. Il est en général assez difficile, en effet, d'isoler dans les processus éducatifs des pratiques d'enseignement de nature à en déduire en toute rigueur une causalité



entre telle façon de faire et tel progrès de l'apprentissage chez les élèves. La complexité de ce qui passe dans les interactions humaines au sein d'une classe est très difficile à ramener à quelques variables qu'on peut tester à souhait (Rey, 2014). Résumer par exemple l'évaluation formative à la question du *feedback* risque d'appauvrir significativement cette approche sous prétexte d'identifier une technique mesurable dans les canons expérimentaux de la psychologie cognitive.

Dans une revue de littérature critique, Bennett (2011) relève d'ailleurs la fragilité des relations causales invoquées par les avocats de l'évaluation formative dès lors qu'on souhaite utiliser les techniques des méta-analyses pour les expériences considérées. Shepard (2005) comme Shavelson (2008) estiment d'ailleurs qu'une condition nécessaire de réussite de l'évaluation formative réside dans son intégration au cœur des curricula d'enseignement, pour être déclinée en termes spécifiques adaptés aux différentes disciplines. Autrement dit, il est non seulement difficile d'évaluer les effets propres de l'évaluation formative « à système éducatif égal par ailleurs », mais une condition même de l'efficacité d'une approche en termes d'évaluation formative réside dans sa diffusion au sein des différents enseignements disciplinaires qui la rend forcément moins visible et isolable.

## ÉVALUER ET ENSEIGNER

[Les pratiques d'évaluation des enseignants au collège \(enquête DEPP 2004\)](#)



### Un jugement sous influences : les effets des relations et des représentations

L'analyse psychologique de l'évaluation amène à s'interroger sur les relations maître-élève et leurs conséquences sur l'évaluation de l'élève et par rebond sur ses apprentissages (et inversement). Les chercheurs en psychologie sociale se réfèrent aux malentendus sociocognitifs (qui vont détourner l'élève des apprentissages attendus), aux effets des comparaisons sociales (sentiment d'injustice, estime

de soi, confiance), au style d'autorité de l'enseignant (motivation, engagement, ou retrait), etc.

L'évaluation est puissamment influencée par le passé scolaire ou familial de l'élève : « *les notes antérieures sont de meilleurs prédicteurs que les résultats à des tests objectifs* » (De Ketele, 1993).

De nombreux travaux mentionnent l'influence sur l'évaluation et les apprentissages des attentes de l'enseignant envers les élèves : « *les attentes joueraient le rôle de filtres interprétatifs, susceptibles de conduire à des distorsions de la réalité lorsque l'enseignant perçoit, interprète et évalue les actions d'un élève* » (Trouilloud & Sarrazin, 2011). L'enseignant privilégie implicitement des informations qui coïncident avec ses attentes. Il va juger *a priori*, évaluant plus favorablement les élèves pour lesquels ses attentes en termes d'effort étaient plus fortes. On constate également que cette attitude bienveillante envers certains peut être influencée par des stéréotypes (genre, minorité ethnique, milieu social, etc.).

Lorsque les chercheurs en sciences sociales, en psychologie sociale et en sciences de l'éducation croisent leurs réflexions, ils mettent en évidence ce jugement des enseignants sous influence, ou en tout cas clairement multidéterminé. On parle alors de distorsions cognitives ou de contraintes motivationnelles. Plus simplement, plusieurs biais ont été observés et décrits : l'effet de stéréotypie (un élève bon en maths sera supposé bon en français), l'effet de halo (caractéristiques de l'élève), l'effet de contexte (on juge un élève à l'aune des autres élèves de la classe, autrement dit plus sévèrement dans une classe « forte »), l'effet de correction (fluctuation de la note en fonction du nombre de copie déjà corrigées), l'effet de tendance centrale (notation autour d'un intervalle établi, qui pourrait se rapprocher de la constante macabre) ou encore l'effet de contamination (influence des résultats antérieurs de l'élève, voir De Lansheere, 1974 ; Bonniol & Vial, 2009 ; Castincaud & Zakhartchouk, 2014).

À ces effets s'ajoutent la perception et le respect de normes scolaires qui peuvent influencer le jugement. « *L'adéquation des*



élèves avec les normes sociales est un paramètre du jugement sur lesquels les élèves peuvent jouer » (Dompnier et al., 2011). Pour réduire les inégalités face à l'évaluation, il serait intéressant de donner à tous les élèves une « clairvoyance normative » et à tous les enseignants la clairvoyance des facteurs d'impact de leur jugement : « Sensibiliser les enseignants sur les conséquences comportementales et perceptives de leurs attentes pourrait constituer un levier intéressant pour réduire l'amplitude » de ce biais (Trouilloud & Sarrazin, 2011).

### Se professionnaliser pour évaluer

« Le jugement professionnel en évaluation n'est pas un jugement mécanique consistant à "mesurer un écart" entre les objectifs définis et la performance évaluée » (Cooksey et al., 2007). C'est un acte de discernement, une « capacité à construire une intelligibilité des phénomènes d'évaluation en situation » (Tessaro, 2013). Le jugement professionnel participe d'une éthique de la responsabilité (professionnalité) : « l'enseignant envisage les conséquences pratiques et morales de ses actions » (Mottier Lopez & Allal, 2010, reprenant Weber, 1919).

Assez paradoxalement, les auteurs ayant étudié la modélisation de l'évaluation (Cardinet, Bonniol & Vial), restent très circonspects, voire hostiles à l'idée de juger les élèves. Cela rejoint également des thèses nord-américaines, qui abordent l'évaluation scolaire d'un point de vue moral (voire religieux : « Tu ne jugeras point ») : l'évaluation doit faire partie d'une démarche bienveillante, d'accompagnement, de confiance, en refusant tout classement, tri et étiquetage.

Pour autant, on peut considérer que le jugement de l'enseignant peut être « professionnel », en ce sens que l'enseignant doit savoir « maîtriser les "techniques de construction des contrôles écrits à référence critériée" (formulation des items, objectifs, niveaux d'exigence, types de situations, cotation, pondération, barème, etc.) » (Tessaro, 2013).

Pour ce faire, Mottier Lopez et Allal (2010) avancent l'idée d'une triangulation pour mettre en relation des données issues de plusieurs méthodologies ou sources. Ainsi, deux maîtres mots sont à retenir, confronter et collaborer :

- confronter les résultats de plusieurs contrôles ;
- comparer les résultats de différentes modalités d'évaluation (contrôle écrit, productions, traces d'activités, observations, etc.) ;
- prendre l'avis de collègues connaissant l'élève ;
- se référer à des savoirs théoriques.

Dans le jugement professionnel, intervient la capacité à traiter les informations recueillies lors des phases évaluatives, notamment évaluation diagnostique et évaluation formative. Ce processus suppose « la justification du choix des moyens en lien avec les visées ou intentions et le partage des résultats de la démarche dans une perspective de régulation » (Lafortune & Allal, 2008).

Allal et Mottier Lopez (2009) définissent les trois démarches qui constituent ce processus :

- mise en relation de plusieurs sources d'informations ;
- interprétation de la signification des informations au regard de référentiels de différentes natures ;
- anticipation et appréciation des conséquences probables de plusieurs actions envisagées (idée d'une « éthique de la responsabilité » selon Weber).

Dans ce modèle, où le jugement professionnel combine une perspective de validation et une perspective interprétative, il est nécessaire de croiser le quantitatif et le contextuel, à charge pour l'enseignant de savoir interpréter les écarts entre les informations fournies par des indicateurs chiffrés et une appréhension locale et circonstanciée (Mottier Lopez, 2013).

Là encore, faut-il former les enseignants à être professionnels dans leur jugement ? Si les représentations des enseignants en activité sur l'évaluation font l'objet d'une littérature importante, peu d'études ont été réalisées auprès



des enseignants en formation. Une enquête a été menée en 2010-2011 auprès de 134 enseignants du primaire et 339 aspirants enseignants du canton de Genève. L'enquête était structurée autour de trois ensembles d'items :

- l'évaluation est un processus formatif visant la régulation des apprentissages et la remédiation des difficultés scolaires (5 items) ;
- l'évaluation est un processus visant à appréhender si les élèves ont atteint les objectifs fixés (critères, 5 items);
- les finalités de l'évaluation sommative et normative (4 items).

Il apparaît que les enseignants en formation différencient bien évaluations formative (plus importante), normative et critériée (moins importante), alors que les enseignants en activité ne différencient pas évaluation formative et évaluation critériée, les considérant comme complémentaires. Ces derniers sont par ailleurs plus en accord avec l'énoncé d'une évaluation sanction-sélection. Faut-il en déduire que les enseignants en activité sont plus « traditionnels » ou, comme le suggèrent les auteurs de cette enquête, les enseignants en formation ne se sont pas encore assez frottés « au terrain » et aux contraintes contextuelles (Boraita & Issaieva, 2013) ?

Dans le cadre du même projet de recherche, une autre équipe s'est intéressée à 216 enseignants bulgares, sur les mêmes items. Là où la première recherche vise à établir des différences de conceptions, cette deuxième enquête cherche à voir comment se situent les enseignants bulgares par rapport aux conceptions dominantes du dispositif évaluatif mis en place. Il semble que ceux-ci privilégient la « *fonction sommative dans une perspective critériée et normative* ». Pour les auteurs, cette attitude est la résultante d'une pratique évaluative héritée de la pédagogie soviétique où « *l'évaluation devait tenir compte non seulement des résultats et du succès personnel de l'élève, mais aussi de sa contribution aux résultats du groupe* », malgré une ordonnance de 2009 qui stipule que l'évaluation scolaire est définie comme « *un processus qui permet de voir dans quelle mesure*

*les objectifs de l'enseignement sont atteints sur la base des performances des élèves et des notations* » (Petkova, 2013).

### Construire une culture professionnelle de l'évaluation

Comment développer les compétences professionnelles des enseignants pour une évaluation fiable et équitable ? Là encore il s'agit d'une confrontation du jugement de l'enseignant par rapport aux représentations communes des attentes officielles afin de parvenir à un consensus sur l'appréciation des travaux d'élèves. L'enseignant doit confronter ses jugements avec ses propres attentes, les critères d'évaluation qu'il a construit au fil de son expérience, ses connaissances des élèves (leurs parcours scolaires, leurs caractéristiques). Devant la difficulté à assurer l'adéquation entre activités didactiques et attentes institutionnelles mais aussi à partager une « *compréhension commune de ce qui est attendu du point de vue des apprentissages des élèves* » (Tessaro, 2013), le canton de Genève a introduit dans le parcours de formation continue des enseignants les pratiques de modération sociale.

Il s'agit ici d'appréhender ce que l'équipe de l'université de Genève, dirigée par Mottier Lopez, décrit comme la nécessité de négocier au sein d'un groupe social (équipe enseignante en l'occurrence) des consensus en matière de pratiques évaluatives. Toute démarche évaluative qui se voudrait (et se devrait) collaborative nécessite de clarifier ses pratiques d'évaluation par rapport à celles des autres. Les pratiques peuvent être partagées par le groupe, socialement admises (on admet des divergences de vue eu égard aux prescriptions, à la culture professionnelle ou à sa déontologie) mais peuvent être contestées (mais non empêchées). Cette culture de l'évaluation s'apprend, de même que doivent être envisagés (admis) les arrangements évaluatifs (Merle, 2012b) liés à la représentation qu'a l'enseignant de l'élève (stéréotypes, biographie).

## LES PARADIGMES DE L'ÉVALUATION

En guise de point d'étape et en s'appuyant sur des réflexions plus anciennes de De Ketele (1993), on peut synthétiser quelques-uns des paradigmes de l'évaluation :

- le « paradigme de l'intuition pragmatique », montrant combien l'évaluation est étroitement liée à l'évaluateur (légitimité) et aux exigences de contexte ;
- le « paradigme docimologique » qui, sous couvert de fiabilité, réduit l'évaluation à sa forme sommative et ne permet pas de dégager des objectifs pédagogiques ;
- le « paradigme sociologique » qui, en cherchant l'égalité, ignore les inégalités culturelles et les renforce ;
- le « paradigme de l'évaluation centré sur les objectifs » (pédagogie par objectifs, pédagogie de maîtrise). Les critiques tiennent à la fois à une prescription descendante de l'usage d'outils « validés » et à une confrontation de fait entre enseignants et élèves, les uns jugeant les autres, de même qu'à un oubli des valeurs fondamentales mais non opérationnalisables ;
- le « paradigme de l'évaluation formative dans un enseignement différencié », qui ne répond pas à toutes les questions à se poser en termes d'évaluation pour les apprentissages : faut-il varier les objectifs (baisser le niveau d'exigence), moduler selon les types de compétences (privilégier les compétences de base), différencier ses pratiques ? Faut-il intégrer des éléments tenant à la personnalité de l'enseignant-e ?

Ces différents paradigmes ne se retrouvent-ils pas dans la vision politique de l'évaluation ? Stobart (2011) a analysé le processus d'appropriation et de généralisation des expérimentations en matière d'évaluation formative, notamment en Angleterre et en Écosse. Comme pour toutes autres expérimentations, le soutien politique peut apporter « la légitimité, des ressources supplémentaires, le soutien des cadres intermédiaires et le développement d'une

*expertise locale* » (ce dernier point ayant d'autant plus de poids dans le contexte concurrentiel des écoles anglaises). Il peut *a contrario* induire des malentendus, des blocages, de la méfiance de la part des enseignants, d'autant plus si « cette expérimentation s'accompagne d'exigences bureaucratiques » et si la formation à l'évaluation devient trop « formatée ».

Évaluation diagnostique, formative, sommative, quand est-il opportun de les utiliser, dans quels contextes, pour quels objectifs ? Les bons conseils ne manquent pas pour ne pas sanctionner, ne pas stresser les élèves, pour se fixer des objectifs de réussite des élèves et s'en approcher au plus près, pour être bienveillant et attentif. Comment formaliser ces bonnes recettes qui, pour reprendre les propos de F. Muller, « professionnalisent l'évaluation » ?

« L'absence de différenciation entre évaluation formative et sommative est récurrente. Le mélange des deux empêche les enseignants de connaître le niveau des acquis des élèves et donne aux élèves le sentiment d'être en contrôle permanent, ce qui engendre une forme de "stress" pour certains. Tout ceci est renforcé par la place très réduite des formes d'évaluation reçues positivement par l'élève (notamment l'auto-évaluation) » (Houchot et al., 2013).

## ÉVALUER EN PRATIQUES

De même qu'il existe peu d'études spécifiques sur la façon dont se déroulent concrètement les évaluations, la formation des enseignants dans le domaine est assez modeste, malgré la place centrale de cette activité dans les pratiques scolaires. Si les examens nationaux sont normés, les autres types d'évaluation sont souvent laissés à l'initiative des enseignants, qui se débrouillent entre échanges de pratiques plus ou



moins organisées, expérimentations et bricolage dans la classe.

Dans ce contexte, une hybridation des pratiques caractérise les évaluations formatives ou les évaluations formatrices, les acteurs éducatifs jonglant avec des principes théoriques divers et devant concilier une réelle liberté pédagogique en cours de formation avec les cadres réglementaires (bulletins, examens, diplômes...) et les habitudes sociales existantes (échelles de notation par exemple).

En France, la base de données [Expéri-thèque](#), qui répertorie les expérimentations pédagogiques, permet d'avoir un panorama des « innovations » en matière d'évaluation. 314 fiches concernent une démarche d'auto-évaluation, 957 l'évaluation par compétences (partielle ou non), 99 l'évaluation par les pairs, 52 le portfolio et 142 la suppression de la note. Ces innovations font écho aux réflexions actuelles sur les pratiques censées améliorer l'évaluation pour apprendre.

## AUTO-ÉVALUATION ET ÉVALUATION PAR LES PAIRS

L'une des pistes privilégiées consiste à briser le lien descendant et hiérarchique de l'évaluation du maître vers l'élève, pour rendre ce dernier acteur de son évaluation. Dans une perspective d'engagement des élèves, « *l'auto-évaluation, par conséquent, n'est pas qu'une pratique d'évaluation ; c'est aussi une activité d'apprentissage. C'est une manière d'encourager les élèves à réfléchir sur ce qu'ils ont appris, à chercher les moyens d'améliorer leur apprentissage, et à planifier ce qui leur permettra de progresser en tant qu'apprenants et d'atteindre leurs objectifs. [...] En tant que telle, elle comprend des compétences en termes de gestion du temps, de négociation, de communication – avec les enseignants et avec les pairs – et d'autodiscipline, en plus de la réflexivité, de l'esprit critique et de l'évaluation* » (Broadfoot, 2007).

Il existe souvent une confusion entre auto-évaluation, qui implique une réflexion sur son apprentissage, et la simple auto-notation, qui peut rester à un stade d'engagement formel mais superficiel des élèves dans leur apprentissage (Nusche *et al.*, 2013).

En tout état de cause, l'interdépendance entre apprentissage et évaluation rend d'autant plus cruciale leur planification conjointe, et plaide en faveur d'une plus forte implication des élèves dans le processus évaluatif, en lien avec le développement d'une approche formative (CERI, 2005). Cette dimension participative peut prendre plusieurs formes, utiliser différentes techniques et intervenir à différents moments pédagogiques. Certaines pratiques régulières peuvent toutefois être observées (CERI, 2005 ; Stobart, 2011) :

- recours à l'étayage des apprentissages : l'enseignant ne fournit pas de réponse, mais suggère des pistes pour encourager les élèves à poursuivre leur apprentissage ;
- constitution d'un répertoire de stratégies d'apprentissage en lien avec la manipulation de concepts complexes, à l'aide par exemple de cartes conceptuelles, pour permettre aux élèves de faire le point sur les connaissances déjà acquises et prioriser celles à acquérir ;
- développement de compétences d'auto-évaluation, nécessitant à un moment ou à un autre une confrontation entre les appréciations des élèves et celles de l'enseignant ;
- développement du rôle des élèves dans l'évaluation entre pairs : promotion de la critique constructive à l'aide de grilles critériées par exemple, et du tutorat entre élèves à l'aide de rubriques (outils de notation associant une échelle de points à des critères de qualité).

L'auto-évaluation est en effet souvent liée avec l'évaluation collective ou coopérative. D'après Black et Wiliam, l'évaluation entre pairs peut favoriser la construction de cette vue d'ensemble en offrant un cadre stimulant pour introduire l'auto-évaluation : elle « *contribue à dynamiser le cadre d'apprentissage, aide les élèves à développer leurs aptitudes sociales, et*

les prépare à l'auto-évaluation » (CERI, 2005). Les interactions verbales entre élèves autorisent des critiques différentes de celles de l'enseignant et formulées dans une langue naturellement plus usuelle, même si les élèves jouent à endosser le rôle de l'enseignant. Les élèves peuvent plus facilement identifier les obstacles et les solutions en suivant le raisonnement de leurs pairs que celui de l'enseignant. Il semble que les élèves s'approprient plus facilement les critères d'évaluation en examinant un travail qui n'est pas le leur.

Dans une perspective d'évaluation des compétences du XXI<sup>e</sup> siècle, le travail coopératif est, faut-il le rappeler, une dimension essentielle : susciter des activités collectives dans un domaine aussi crucial est une façon d'habituer les élèves à travailler ensemble dans un secteur qui semble spontanément marqué par la compétition individuelle (Pepper, 2013).

Enfin, le travail entre pairs constitue un vecteur d'introduction d'éléments de différenciation pédagogique dans le contexte de la classe : les élèves qui ont compris un problème peuvent tenter de l'expliquer avec d'autres qui rencontrent des difficultés, pendant que l'enseignant se concentre sur des blocages particuliers.

Les compétences sociales indispensables au bon déroulement d'une activité d'évaluation entre pairs peuvent cependant ne pas être naturellement convoquées à bon escient. Autrement dit, pour être efficace, l'évaluation entre pairs doit bénéficier d'un guidage, voire d'un apprentissage. Une des activités fréquemment mentionnée est celle de la notation des devoirs à la maison à l'aide de feux tricolores : chaque élève dessine sur sa propre copie le feu correspondant au degré de confiance qu'il accorde à sa performance. En classe, les élèves qui ont dessiné un feu vert ou orange travaillent ensemble pour s'évaluer ou s'entraider, tandis que l'enseignant porte son attention sur ceux qui sont les moins sûrs de leur performance (feu rouge).

Mais globalement, les « preuves » en faveur de l'impact positif de l'évaluation entre pairs sont peu abondantes ; le fait que ces pratiques soient ou non répan-

dues (formellement) n'est d'ailleurs pas non plus attesté. Examinant le tutorat entre pairs, Baudrit a constaté que les régulations les plus élevées mises en œuvre par les élèves semblent le fruit d'une certaine asymétrie entre le tuteur et le tutoré (Baudrit, in Allal & Mottier Lopez, 2007). Autrement dit, il faut un écart suffisant pour justifier une relation de complémentarité qui assure une vraie confrontation de démarches d'apprentissages et évite le flou et l'incertitude dans lesquels risquent d'être plongés les pairs en dyades (ou paires) trop homogènes. Inversement, un trop grand écart risque de faire basculer la relation dans une situation d'expertise, avec un rôle trop passif du tutoré, moins propice à un apprentissage.

[Une dictée particulière](#)



### Les biais de l'auto-évaluation : le poids de la comparaison sociale

Tout comme le jugement de l'enseignant subit un effet de contexte lié au niveau de la classe, la qualité de l'auto-évaluation des élèves sera infléchie par le niveau de performance globale de la classe ou de l'établissement : les élèves se sentent moins performants dans des établissements sélectifs. Ces observations ont produit le concept de soi scolaire et l'effet de contraste ; « *il vaut mieux être un gros poisson dans une petite mare* » que l'inverse (Davis, 1966, cité par Dumas & Huguet, 2011). Les élèves ont par ailleurs tendance à se comparer à l'intérieur de leur groupe de sexe ou à évaluer leur performance en fonction d'un camarade d'un niveau légèrement supérieur.

### DU TEMPS POUR BIEN UTILISER LES « ERREURS »

Pour l'enseignant, l'évaluation renvoie à des choix sur ce qu'il est important de faire apprendre et la manière de le faire. Les exercices d'évaluation doivent permettre à l'élève de comprendre où il en est dans ses apprentissages, par rapport aux objectifs qui lui ont été assignés. Pour l'enseignant, cela suppose de ne pas se focaliser sur les erreurs et le besoin de



rendre des comptes hors de la classe, mais bien d'appréhender les écueils que l'élève à rencontrer lors de l'exercice et les remédiations à envisager.

Comme le montrent les expérimentations conduites par Black et William, la phase de dialogue ou de questionnement, constitutive d'un « bon » processus d'évaluation formative montre l'importance de laisser du temps aux élèves pour qu'ils questionnent, réfléchissent et formulent au mieux leurs réponses, plutôt que des réponses convenues et attendues par l'enseignant (Muller & Normand, 2013). De même, l'enseignant prend le temps de (re)formuler ses questions, ses réponses et les relances nécessaires à une meilleure compréhension des élèves. Dans les années 1970, des recherches orientées plus spécifiquement sur les échanges à caractère évaluatif (*instructional dialogues, assessment conversations*) ont conceptualisé ces phases d'interactions (quelles stratégies, quel langage, quelle compétences de communication, etc.). Pour être tout à fait efficaces, ces moments d'échanges évaluatifs doivent être guidés par un objectif commun ; apprendre et faire apprendre. Pour ce faire, le dialogue doit être réellement interactif et dialogique et ces phases de questions-réponses évaluatives doivent devenir des outils d'étayage de l'enseignement tout en enrichissant les compétences sociocognitives indispensables aux élèves (c'est aussi par ce dialogue que passent les consignes de travail, les explicitations quant aux objectifs de la tâche, voir Ruiz-Primo, 2011).

Dans son travail sur la place de l'erreur à l'école, Reuter (2013) propose de l'intégrer de façon explicite dans la situation didactique, afin de pouvoir surmonter cette erreur et de s'en servir comme levier d'enseignement dans certains dispositifs. Pour lui, l'erreur est en effet un dysfonctionnement inhérent à la situation d'apprentissage, qu'il convient par conséquent de prendre en charge autrement que par la stigmatisation ou par la répétition (refaire un exercice jusqu'à ce que l'erreur disparaisse). L'erreur n'est pas dans cette perspective une faute (au sens « honteux » du terme) mais un problème qui peut avoir diverses causes (dans la classe ou hors de la classe) qu'il convient d'identifier autant que possible pour lui apporter des réponses adaptées.

Il s'agit, à rebours d'une tradition consistant à cacher ses erreurs par divers procédés (gomme, effacement ou abstention en cas de doute), de les intégrer dans le cadre d'une évaluation qui ne peut être que formative. En dernière instance, l'objectif reste bien évidemment de résoudre les difficultés et de dépasser les obstacles, pas de donner un caractère naïvement « positif » à l'erreur, mais en comprenant qu'on apprend avec ses erreurs, c'est-à-dire en les identifiant pour pouvoir se donner les moyens de les comprendre et de les éviter lors de l'évaluation certificative qui interviendra à un moment ou un autre.

## **COMPÉTENCES ET PROJETS : COMMENT ÉVALUER LA MOBILISATION ET L'INTÉGRATION DES RESSOURCES ?**

Depuis la fin des années 2000, des chercheurs, des militants pédagogiques et des responsables de politiques éducatives ont utilisé le terme de « compétences » comme levier d'évolution des pratiques et des systèmes éducatifs (Rey, 2011).

Ils expriment là en général leur souci de mieux travailler sur le transfert des apprentissages et de s'assurer que les élèves savent mobiliser et intégrer des ressources de diverses natures pour répondre à la complexité des situations que l'on peut rencontrer dans la vie.

Contrairement aux tâches automatisables que l'on qualifie parfois de compétences dans les situations scolaires ordinaires (effectuer une multiplication, accorder le verbe avec le sujet, utiliser un fond de carte...), ces compétences-là renvoient à une situation originale et complexe (par exemple, écrire un texte en tenant compte du destinataire et de l'effet recherché) pour laquelle il s'agit non seulement de mobiliser des connaissances mais aussi d'interpréter correctement la situation pour choisir les connaissances pertinentes à mobiliser.

L'évaluation de ces compétences représente donc une rupture importante avec les modes traditionnels d'évaluation, puisqu'elle implique que l'on évalue à travers des situations complexes et que l'on

privilégie la compréhension globale d'une réponse singulière sur la mesure de résultats standardisés (Pepper, 2013).

Dans la plupart des cas, il n'y a effectivement pas une seule bonne réponse mais plusieurs stratégies possibles. En outre, divers travaux de recherche ont montré que ce que B. Rey (2014) appelle le « regard instruit », c'est-à-dire l'interprétation savante de la situation qu'attend l'école, n'a rien d'évident. Plus on se rapproche de situations dites « de la vie réelle », plus l'élève peut être tenté de privilégier une efficacité spontanée par rapport aux objectifs affichés de la situation, sans forcément recourir à un cadrage « scolaire » de la situation. Comment concilier évaluation authentique des compétences et mobilisation par l'élève de catégories issues des savoirs ?

La difficulté peut effrayer et conduire à refuser la perspective des compétences au nom d'un caractère plus « démocratique » des évaluations traditionnelles, qui mettent en jeu des savoirs scolaires dont la répétition intensive pourrait garantir la réussite, au sein du cadre rassurant et éprouvé des disciplines et dans un contexte explicitement scolaire. C'est en même temps éviter de se confronter à la question de la particularité du regard scolaire sur les choses et de la relation d'un individu à la spécificité de la culture de l'école.

Une autre réponse consiste à préparer de façon intensive les élèves au transfert impliqué par l'approche en termes de compétences, et à leur faire travailler en particulier de façon explicite l'axe de la complexité tout autant que celui de la difficulté (Roegiers, 2014). L'école met traditionnellement au centre de l'attention la difficulté, c'est-à-dire le passage du connu au pas encore connu, mais rarement la complexité, c'est-à-dire les nouvelles articulations de ressources déjà connues.

On comprend que l'évaluation des compétences, par sa nouveauté et sa complexité de mise en œuvre, se prête particulièrement mieux à une évaluation formative qu'à des évaluations sommatives ou certificatives. La mise en œuvre du livret personnel de compétences en France a d'ailleurs montré toute la difficulté de vouloir faire coexister une logique de certification standardisée et une logique de certifica-

tion des compétences, aggravée par le maintien d'un examen traditionnel comme le brevet (Hazard *et al.*, 2012).

La nature même des compétences qu'on cherche à évaluer est peu compatible avec les évaluations standardisées à grande échelle qui peuvent difficilement appréhender la performance des élèves en matière de résolution de problèmes, de raisonnement ou de travail collaboratif. Par voie de conséquence, ces évaluations procurent peu d'informations détaillées de nature à diagnostiquer les difficultés particulières rencontrées. En fait, une évaluation certificative des compétences renvoie à un modèle d'évaluation qualitative qui a peu à voir avec le modèle traditionnel de la mesure. L'importance de la dimension qualitative implique que l'activité évaluative est plus centrale que le dispositif lui-même et l'évaluation repose d'abord sur la confiance qu'on peut accorder à des évaluateurs qui doivent être vraiment formés à cette mission (Bain, 2014).

On retrouve des problématiques voisines dans les cas d'apprentissage par projet, pour lesquels les principes les plus adaptés (auto-évaluation, évaluation par les pairs, portfolios, critères explicites co-construits par les élèves...) relèvent plutôt de l'évaluation formative, voire formatrice (Reverdy, 2013).

## DES CLASSES SANS NOTES

Si les notes sont d'autant plus présentes que l'on progresse dans les strates du système éducatif, des enquêtes de l'Inspection générale sur les pratiques d'évaluation à l'école primaire ont mis en relief une évolution importante à ce niveau : les appréciations littérales et les validations des acquis sont plus souvent utilisées que les notes (IGEN, 2013). Codes couleurs, lettres, *smileys*, ceintures de judo : diverses alternatives permettent de compléter les appréciations littérales et d'indiquer où se situe l'élève dans les apprentissages, sans recourir aux notes.

On retrouve au sein même de l'école primaire des tendances récurrentes dans tout le système : plus on s'élève dans les cycles, plus on recourt aux notes, la pression du collège proche s'exerçant particulièrement sur les classes de CM1 et CM2.



Par ailleurs, ce sont les matières dites « fondamentales », perçues de l'ordre des « connaissances », qui attirent la note (français, mathématiques, voire histoire et sciences) pendant que tout ce qui relève des domaines artistiques, sportifs, des compétences sociales ou civiques est plus rarement redevable d'une évaluation chiffrée. Il reste que ces modes d'évaluation ne sont pas toujours mis en cohérence au niveau de l'établissement, chaque enseignant privilégiant parfois une démarche personnelle au nom de la liberté pédagogique.

### Les expérimentations de classes sans note

La base de connaissances [Expérithèque](#) offre quelques exemples de classes sans note. Bien souvent l'expérimentation commence par une seule classe par collège ou un seul niveau (83 % en classe de 6<sup>e</sup>) et dans des collèges ayant des élèves en difficultés. Dans quasiment toutes les situations, ce sont l'évaluation par compétences et le LPC qui signent l'arrêt de la notation. Les objectifs et les résultats déclarés montrent une plus grande concertation entre enseignants, une meilleure communication avec les familles, la lutte contre le décrochage et l'absentéisme, l'amélioration du climat scolaire, la remotivation des élèves, l'amélioration des résultats scolaires, un prise en charge différenciée, etc.

Les enseignants expérimentateurs citent en positif les ressources ou points d'appui tels que « *la formation des enseignants à l'évaluation par compétences ; le travail en lien avec le CERCA à poursuivre (questionnaires auprès de tous les élèves de 6<sup>e</sup> concernés) ; les échanges avec certains IPR et toutes les lectures et investigations effectuées pour nourrir notre réflexion et nos arguments* ». Quant aux principales difficultés rencontrées, les enseignants font part des « *programmes dans certaines disciplines (lettres, histoire géo) [qui] n'ont pas été pensés en fonction des compétences ; le temps de concertation très important et nécessaire [qui] fait parfois défaut ; l'intégration des nouveaux collègues ; [la nécessité de] lutter contre les "croyances naïves" et "le formatage de la note"* ».

Les expériences de classes sans note dans l'académie de Poitiers  
[Illustration de classes sans note : une expérience dans l'académie de Poitiers](#)



## LE NUMÉRIQUE, VECTEUR DE TRANSFORMATION DE L'ÉVALUATION ?

Beaucoup d'attentes sont placées dans le développement des technologies numériques comme facteur d'amélioration des procédures d'évaluation.

### L'évaluation aidée par le numérique

De nombreuses innovations ont été expérimentées en la matière depuis une quinzaine d'années et un marché, avec ses enjeux économiques et politiques, existe même dans certains secteurs particuliers. Ainsi, dans des pays où les établissements et les autorités éducatives locales achètent leurs dispositifs d'examen, de nombreux efforts sont entrepris par des entreprises pour proposer des batteries de tests de plus en plus performants et sophistiqués. L'intégration dans ces outils d'un certain état de l'art de la recherche en matière d'évaluation est même utilisée comme argument commercial pour valoriser leurs produits. Néanmoins, les auteurs d'une revue de littérature anglophone en « e-assessment » constatent que le numérique est souvent utilisé pour renforcer les méthodes d'évaluation traditionnelles ou que les transformations sont cantonnées dans des poches d'innovation qui ne se diffusent pas dans le système éducatif (Timmis *et al.*, 2012).

Si le numérique se contente de permettre des tests massifs sur écran ou des QCM en ligne, comme cela se remarque fréquemment, la plus-value est en effet modeste. Bien que plus rares, des dispositifs permettant une adaptation des questions en fonction des réponses précédemment formulées montrent comment des évaluations peuvent devenir plus interactives et enrichies pas les retours des élèves en



temps réel. En la matière, certaines innovations se développent plus rapidement dans le cadre de l'enseignement supérieur que dans l'enseignement scolaire.

Globalement, plusieurs bénéfices peuvent être attendus d'un recours au numérique pour véritablement transformer l'évaluation :

- permettre un retour immédiat vers l'apprenant pour infléchir précocement l'apprentissage ;
- faciliter un apprentissage auto-régulé et personnalisable ;
- aider un apprentissage collaboratif (évaluation par les pairs, partage d'activités, interactions...);
- fournir des situations d'évaluation plus authentiques (situations complexes, résolution de problèmes, hypothèses à tester...);
- diversifier les données évaluables (compétences, connaissances, traces cognitives et tâches prises en compte) ;
- fournir des réponses plus flexibles et appropriées (en termes de rythmes et de lieux d'apprentissage) ;
- rendre les corrections plus efficaces et réduire la charge de travail des enseignants ;
- rendre l'expérience d'évaluation plus riche et plus intéressante pour les élèves ;
- améliorer les performances des élèves, notamment grâce à un meilleur engagement dans les activités ;
- articuler mieux les dimensions formatives et sommatives des évaluations (suivi en continu des traces) ;
- améliorer la validité et la fiabilité (données plus fines et plus riches).

### Le Danemark, à la pointe du e-assessment

Le Danemark est souvent cité dans les rapports officiels comme un pays d'avant-garde dans le recours des technologies numériques en matière d'évaluation. Les tests nationaux mis en place en 2008 pour suivre les acquis des élèves tout au long de leur scolarité se déroulent ainsi par ordinateur et sont adaptatifs : les questions proposées s'adaptent aux réponses des élèves. Une réponse incorrecte amène une question plus facile, et une réponse

correcte, une question plus difficile. Strictement canalisée dans le processus de formation, ces tests sont soigneusement conçus pour être utilisés aux fins de palmarès ou de classements publics. Il n'est pas étonnant de constater d'ailleurs que c'est dans un pays à faible pression évaluative que l'on réussit à imaginer des formes plus modernes d'évaluation.

L'utilisation des TIC est d'ailleurs centrale dans les pratiques d'enseignement et d'apprentissage au Danemark, pour assurer l'alignement entre les pratiques, les objectifs pédagogiques et les modes d'évaluation. Dès 1997, l'ordinateur était autorisé dans les examens écrits des voies commerciales et techniques, puis généralisé en 2000 dans toutes les filières (98 % des élèves utilisaient alors déjà un PC dans les examens écrits de la voie commerciale). Depuis 2001, des sujets sont donnés via CD-Rom, les devoirs écrits peuvent être réalisés sur traitement de texte (ou Excel) et la copie à rendre imprimée.

Enfin, depuis 2008, l'usage d'Internet se répand au niveau des examens avec, en 2010, les premiers sujets permettant l'accès en ligne dans certaines matières (danais, mathématiques, sciences sociales, économie...) et, à partir de 2014, lors des examens d'anglais.

Il pourra s'agir, en l'occurrence, non plus de réaliser une traduction ou un thème, car dans ce cas là le recours aux outils Internet videra l'épreuve de son sens, mais de comparer par exemple la pertinence de trois traductions proposées par les services en ligne et d'étayer son argumentation.

### Le portfolio, outil d'enrichissement de l'évaluation

En dehors du contexte des évaluations sommatives, l'une des promesses du numérique pour l'évaluation est certainement de permettre de faire la preuve de ses talents de façon plus variée qu'à travers les dispositifs « papier-crayon » et sous différents formats. Il revient alors à l'élève de choisir quel est le format le plus adapté (récit, image, vidéo, schéma, tableur, etc.) et le plus créatif pour restituer une compétence. Cette éducation au choix représente d'ailleurs un des éléments éducatifs stratégiques au sein de ce



qu'on appelle communément les « compétences du XXI<sup>e</sup> siècle ». Les potentialités du portfolio numérique ont dans ce contexte particulièrement attiré l'attention.

Le portfolio est un dispositif d'évaluation souvent étudié, utilisé et commenté dans le contexte de l'éducation des adultes et de l'enseignement supérieur. Moins fréquent dans le contexte de l'enseignement scolaire, il est apparu au début des années 2000 comme un outil prometteur pour appréhender des réalisations diversifiées des élèves, dans un format largement plus flexible grâce aux supports numériques.

On peut en retenir la définition de Challis (2005), auteur de travaux sur l'usage des portfolios dans la formation médicale, communément employée au niveau européen : « *des collections sélectives et structurées d'informations, rassemblées avec un objectif précis, et illustrant les réalisations et les progrès [de l'élève], stockées sur un support numérique et gérées par un logiciel approprié ; permettant l'utilisation de fonctionnalités multimédia, intégrées à un site web ou disponible en cd-rom ou dvd-rom* ».

On comprend bien l'alternative que représente le portfolio aux évaluations standardisées et normées : l'élève peut ici montrer les traces d'activités cognitives, de mobilisation des connaissances, de compétences, de réalisations personnelles qui sont plus difficilement prises en compte dans les évaluations habituelles (Pepper, 2013).

La démarche associée à l'usage pédagogique d'un portfolio est souvent décrite dans la littérature sous forme de cycle, faisant ainsi écho à la boucle de rétroaction (*loop*) constitutive du *feedback*. Barrett (2004) par exemple découpe le processus en cinq phases : collecter (*collecting*), sélectionner (*selecting*), réfléchir (*reflecting*), montrer ou partager (*projecting*) et publier ou rendre publiques ses « réussites » (*celebrating*). Graham Attwell propose une grille de lecture, composée de sept fonctions, correspondant à autant de processus pédagogiques différents : identifier (*recognising*), archiver (*recording*), réfléchir (*reflecting*), prouver (*validating*), présenter (*presenting*), prévoir (*planning*), évaluer (*assessing*, voir Attwell et al.,

2007). Malgré ces essais de modélisation, le portfolio numérique, peu visible, semble difficile à appréhender et souffre de discours contradictoires. S'agit-il de privilégier une démarche d'auto-évaluation et de stimuler la réflexivité des élèves ? S'agit-il d'adopter un nouvel outil permettant de valider voire de certifier des acquis ? S'agit-il encore de généraliser l'usage d'un outil standardisé et interopérable ? Les trois discours, formatif, sommatif (voire certificatif) et technique (voire marchand ou politique), semblent opérer dans des espaces relativement cloisonnés, n'offrant guère de prise concrète aux praticiens.

Certaines intentions pédagogiques initiales, qui donnaient la priorité à la dimension réflexive et formative, semblent aussi avoir cédé le pas devant la diffusion d'applications informatiques standardisées privilégiant la production de données d'évaluation traditionnelles (Garrett, 2011).

Scallon (2007) considère que le principal intérêt du portfolio réside dans l'auto-évaluation. Dans cette perspective, il distingue trois types d'usage, pouvant correspondre à autant de jalons dans l'approfondissement progressif d'une démarche d'auto-évaluation :

- le dossier d'apprentissage, qui collecte tous les travaux de l'élève ainsi que ses commentaires réflexifs, et se construit au fur et à mesure ;
- le dossier de présentation, dans lequel l'élève choisit des travaux selon certains critères, par exemple pour montrer une progression ou raconter un épisode particulier de ses apprentissages ;
- le dossier d'évaluation, qui correspond à une sélection des meilleurs travaux de l'élève, pour étayer l'évaluation d'une compétence, à la fin d'une période de formation.

Scallon postule une articulation entre les dimensions formative et sommative de l'apprentissage. Le dossier d'évaluation permet ainsi l'évaluation institutionnelle, tout en rompant radicalement avec la tradition des tests standardisés. L'évaluation du portfolio, par les pairs et/ou par les enseignants, peut de surcroît éprouver la capacité de l'élève à accepter des jugements extérieurs (et donc à s'auto-évaluer).

Dans des contextes de développement des espaces numériques de travail en contexte éducatif d'une part, et des réseaux sociaux d'autre part, on peut se demander si la question des portfolios n'a pas souffert ces dernières années d'un moindre intérêt, certaines de ses fonctions se voyant réinvesties à travers d'autres supports ? En tout état de cause, il est significatif de constater que le dossier consacré au portfolio sur le site ministériel français figure parmi les dossiers archivés qui ne sont plus mis à jour depuis 2009 ([Eduscol](http://Eduscol)).

## L'ÉVALUATION PAR CONTRAT DE CONFIANCE

Peut-on envisager le système d'évaluation par contrat de confiance (EPCC) cher à Antiby, comme un exemple d'évaluation permettant de contrer la « constante macabre », favoriser le processus d'apprentissage ? Antiby propose l'EPCC comme substitut à une évaluation notée dont il dénonce le dysfonctionnement. On peut considérer que l'évaluation mise en cause est une évaluation sommative, pour faire le point sur l'acquisition des apprentissages.

L'alternative proposée est donc la préparation raisonnée de cette phase sommative, son déroulement et son contenu (annoncer ce qui sera demandé pendant le contrôle, préparer le contrôle sous forme de questions-réponses, corriger). Ce contrat va sans doute mieux motiver les élèves, mais pour un objectif bien précis qui est le contrôle noté de leurs connaissances et qui ne prévient pas la répétition de la « constante macabre » (essentiellement liée à la pratique enseignante). Certains, travaillant en conformité avec l'idée de socle commun de compétences et de connaissances, estiment que les bases du contrat de confiance ne peuvent être satisfaisantes. La méthode de l'EPCC considère qu'une compétence sera acquise dès lors que l'élève aura obtenu une note supérieure à 10/20 au contrôle portant sur une liste d'activités associées à la compétence. La question que posent les détracteurs de l'EPCC est donc la suivante : peut-on valider une compétence à partir d'une somme d'items et qui plus est en

se contentant d'un 10/20 ? L'ambiguïté réside plus dans la pertinence de l'évaluation et non dans la manière d'aborder l'acquisition des compétences (la mise en confiance facilitant cette acquisition, voir [Louzach](#) cité par Castincaud & Zakhartchouk, 2014).

D'autres contributions favorables à l'EPCC restent toujours dans ce champ restreint de l'évaluation sommative, y compris dans une approche par compétences, tout en évoquant ponctuellement la nécessité pour les élèves de savoir quelles compétences sont validées et quelles compétences sont encore à travailler (Moisan, 2011). L'EPCC reste ainsi fortement disciplinarisée dans la pratique. On peut par ailleurs retenir l'idée plus globale d'introduire plus d'équité dans l'évaluation en clarifiant le contrat d'évaluation, notamment en circonscrivant de façon claire les contours de ce qui va être évalué (Roegiers, 2014).

## CONCLUSION

En France, le Conseil supérieur des programmes, institué par la loi de refondation de l'école, a commencé ses travaux en octobre 2013. En juin 2014, alors que le terme de curriculum semble trouver difficilement sa légitimité auprès de certains acteurs éducatifs, la réforme de l'évaluation des élèves semble inéluctable. C'est l'évaluation notée, la mise en cohérence de l'évaluation certificative de fin de collège et du livret de compétences (par ailleurs trop complexe ou mal utilisé) qui sont évoquées en premier lieu.

Or, si la refonte de l'évaluation, qu'elle soit certificative ou sommative, est nécessaire, il convient de donner toute sa force à l'ensemble du processus évaluatif pour les apprentissages. Les travaux de recherche consultés dans le cadre de cette revue de littérature montrent la nécessité d'une réflexion sur les modalités de l'évaluation mais aussi sur les motivations de cette évaluation. Il ne s'agit pas d'évaluer des individus mais bien l'état des connaissances et compétences acquises, en conformité avec un parcours curriculaire déterminé. Chaque entité concernée doit y prendre sa part de responsabilités, pour construire ce



parcours continu de formation, à charge pour les enseignants de construire un processus évaluatif garantissant ce parcours et aux cadres de l'éducation de les former et les accompagner en ce sens. Ce tour d'horizon des dispositifs, principes et modalités d'évaluation des élèves ne peut être déconnecté des contenus d'enseignement. L'évaluation n'est pas un but en soi, il s'agit d'évaluer ce qui est acquis et d'évaluer pour apprendre. Mais apprendre quoi ? Il convient donc de s'entendre sur ce que sont les savoirs, connaissances et compétences à apprendre. Comme le stipule la [proposition de socle commun](#) du Conseil supérieur des programmes, « *la validation du socle commun pour tous les élèves est aussi stratégique que les contenus enseignés* ». L'inverse est vrai et la réflexion et la mise en œuvre d'un curriculum tout au long de la scolarité obligatoire doit intégrer, pas à pas mais aussi en continu, les démarches évaluatives.

On note une corrélation positive entre de bons résultats aux évaluations internationales et un cadre opérationnel cohérent pour les évaluations, c'est-à-dire une gestion des contenus articulée à une définition claire de l'évaluation au niveau national et local, ce que l'on appelle aussi l'alignement du curriculum.

*« C'est lorsque les équipes s'interrogent sur la façon d'enseigner et sur la manière "de mieux faire réussir les élèves" qu'elles en arrivent tout naturellement à réfléchir sur la forme de l'évaluation » (Houchot et al., 2013).*

Comme le disaient Black et Wiliam en 2010, l'évaluation formative n'est pas un remède miracle (*magic bullet*), les conditions de réalisation sont trop complexes pour être applicables sans un investissement de tous les acteurs et une impulsion forte des décideurs. Pour les défenseurs de l'évaluation pour les apprentissages, il ne fait pas de doute qu'un changement de culture du « faire apprendre » amènerait une amélioration des standards et un impact positif sur les acquis des élèves. Au-delà d'une préoccupation évaluative, l'école ne pourrait-elle pas s'emparer du concept d'apprendre à apprendre, en favorisant l'évaluation formative plutôt que l'évaluation sommative, en poussant les élèves vers l'autonomie, la maîtrise des connaissances et surtout de leur utilisation ?

## BIBLIOGRAPHIE

Vous retrouverez ces références et quelques autres dans notre [bibliographie collaborative en ligne](#), qui comprend les références complètes et, le cas échéant, des accès aux articles cités (en accès libre ou en accès payant, selon les cas et selon les abonnements électroniques souscrits par votre institution).

- Airasian Peter W. & Russell Michael K. (2008). *Classroom assessment : concepts and applications* [6<sup>e</sup> éd.]. Boston : Burr Ridge.
- Allal Linda (2008). « Évaluation des apprentissages ». In van Zanten Agnès (dir.). *Dictionnaire de l'éducation*. Paris : Presses universitaires de France, p. 311-314.
- Allal Linda & Mottier Lopez Lucie (2005). « L'évaluation formative de l'apprentissage : revue de publications en langue française ». In OCDE, *L'évaluation formative : pour un meilleur apprentissage dans les classes secondaires*. Paris : OCDE, p. 265-290.
- Antibi André (2003). *La constante macabre ou comment a-t-on découragé des générations d'élèves ?* [s.l.] : Math'Adore.
- Assessment Reform Group (2002). *Assessment for learning: 10 principles*. Londres : Assessment Reform Group.
- Attwell Graham, Chrzaszcz Agnieszka, Pallister John, Hornung-Prähauser Veronika, Hilzensauer Wolf (2007). « Grab your future with an E-portfolio : Study on new qualifications and skills needed by teachers and career counsellors to empower young students with the E-portfolio concepts and tools ». *MOSEP Report - More Self-Esteem with my e-Portfolio*. Salzburg : Salzburg Research Forschungsgesellschaft.
- Bain Daniel (2014). « Généralisabilité et évaluation des compétences : pistes et fausses pistes ». In Dierendonck Christophe, Loarer Even, Rey Bernard (dir.), *L'évaluation des compétences en milieu scolaire et en milieu professionnel*. Louvain-la-Neuve : De Boeck, p. 141-166.
- Barbarant Olivier (2014). *Une nouvelle évaluation : construction et expérimentation d'un barème graduel pour l'exercice de la dictée*. Note pour le ministère de l'Éducation nationale (DGESCO).
- Barrère Anne (2009). « Les élèves face au travail scolaire ». In Duru Bellat Marie et van Zanten Agnès (dir.) *Sociologie du système éducatif*. Paris : Presses Universitaires de France.
- Barrère Anne (2014). Conférence au colloque « *Évaluation des élèves et démocratie* », SNP-DEN, ENSMM Besançon, 14 mai.
- Barrett Helen (2004). *Electronic portfolios as digital stories of deep learning. Emerging digital tools to support reflection in learner-centered portfolios*. Puyallup : Electronicportfolio.org.
- Baudrit Alain (2007). « Tutorat entre pairs : Les processus de régulation mis en œuvre par les élèves ». In Allal Linda & Mottier Lopez Lucie (dir.). *Régulation des apprentissages en situation scolaire et en formation*. Bruxelles : De Boeck, p. 113-127.
- Bennett Randy Elliot (2011). « Formative assessment: a critical review ». *Assessment in Education: Principles, Policy & Practice*, vol. 18, n° 1, p. 5-25.
- Black Paul et Wiliam Dylan (1998). *Inside the black box : Raising standards through classroom assessment*. Londres : NFER Nelson.
- Black Paul et Wiliam Dylan (2010). « Inside the black box : raising standards through classroom assessment ». *Kappan Magazine*, vol. 92, n° 1, p. 81-90.
- Bonniol Jean-Jacques & Vial Michel (2009). *Les modèles de l'évaluation : textes fondateurs avec commentaires* [2<sup>e</sup> éd.]. Bruxelles : De Boeck.
- Boraita Fanny, Issaieva Élisabeth (2013). « Les conceptions de l'évaluation chez les futurs enseignants et enseignants en exercice : une étude dans le canton de Genève ». In *Évaluation et autoévaluation : quels espaces de formation ?* Colloque de l'ADMEE, Fribourg, 9-11 janvier.
- Braxmeyer Nicole, Guillaume Jean-Claude, Lévy Jean-François (2005). *Les pratiques d'évaluation des enseignants en collège*. Paris : Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche, Direction de l'évaluation et de la prospective.



- Bressoux Pascal & Pansu Pascal (2003). *Quand les enseignants jugent leurs élèves*. Paris : Presses universitaires de France.
- Broadfoot Patricia (2007). *An Introduction to assessment*. Londres : Continuum.
- Broadfoot Patricia & Pollard Andrew (2006). « The changing discourse of assessment policy: The case of english primary education ». In Lander Hugh, Brown Phillip, Dillabough Jo-Anne & Halsey Albert Henry (dir.). *Education, Globalization & Social Change*. New-York : Oxford University Press.
- Butera Fabrizio, Buchs Céline & Darnon Céline (dir.) (2011). *L'évaluation, une menace ?* Paris : Presses universitaires de France.
- Cardinet Jean (1989). « Évaluer sans juger ». *Revue française de pédagogie*, n° 88, p. 41-52.
- Castincaud Florence, Zakhartchouk Jean-Michel (2014). *L'évaluation, plus juste et plus efficace : comment faire ?* Coll. « Repères pour agir ». Amiens : Canopé-CRDP.
- Centre pour la recherche et l'innovation dans l'enseignement (CERI) (2005). *L'évaluation formative : pour un meilleur apprentissage dans les classes secondaires*. Paris : OCDE.
- Challis Di (2005). « Towards the mature ePortfolio : Some implications for higher education ». *Canadian Journal of Learning and Technology*, vol. 31, n° 3.
- Colardyn Danielle (2009). *Évaluation et examens aux États-Unis : quelles leçons pour l'éducation en Europe ?* Bruxelles : De Boeck.
- Cooksey Ray W., Freebody Peter & and Wyatt-Smith Claire Maree (2007). « Assessment as judgement-in-context: Analysing how teachers evaluate students' writing ». *Educational Research and Evaluation*, vol. 13, n° 5, p. 401-434.
- Crahay Marcel (2007). « Feedback de l'enseignant et apprentissage des élèves : revue critique de la littérature de recherche ». In Allal Linda & Mottier Lopez Lucie, *Régulation des apprentissages en situation scolaire et en formation*. Bruxelles : De Boeck, p. 45-70.
- Crahay Marcel et Mottier Lopez Lucie (2012). *Évaluations en tension*. Bruxelles : De Boeck.
- De Ketele Jean-Marie (1993). « L'évaluation conjugue en paradigmes ». *Revue française de pédagogie*, n° 103, p. 59-80.
- De Ketele Jean-Marie (2013). « L'évaluation de la production écrite ». *Revue française de linguistique appliquée*, vol. XVIII, n° 1, p. 59-74.
- De Landhseere, Gilbert (1974). *Évaluation continue et examens : précis de docimologie*. [3<sup>e</sup> éd.]. Paris : Nathan.
- Del Rey Angélique (2013). *La tyrannie de l'évaluation*. Paris : La Découverte.
- Desclaux Bernard (2014). « Esprit d'entreprise et défiance éducative ». *Administration et éducation*, n° 141, mars, p. 23-29.
- Dompnier Benoît, Pansu Pascal, Bressoux Pascal (2011). « L'évaluation scolaire : une activité multidéterminée ». In Butera Fabrizio, Buchs Céline, Darnon Cécile, *L'évaluation, une menace ?* Paris : Presses universitaires de France, p. 77-84.
- Dumas Florence & Huguet Pascal (2011). « Le double visage de la comparaison sociale à l'école ». In Butera Fabrizio, Buchs Céline, Darnon Cécile, *L'évaluation, une menace ?* Paris : Presses universitaires de France, p. 94-104.
- Endrizzi Laure (2013). « Les lycées, à la croisée de tous les parcours ». *Dossier de veille de l'IFÉ*, n° 88, décembre.
- Endrizzi Laure & Rey Olivier (2008). « L'évaluation au cœur des apprentissages ». *Dossier d'actualité Veille et Analyses*, n° 39, novembre.
- Forquin Jean-Claude (2008). *Sociologie du curriculum*. Rennes : Presses universitaires de Rennes.
- Garrett Nathan (2011) « An e-portfolio design supporting ownership, social learning, and ease of use ». *Educational Technology & Society*, vol. 14, n° 1, p. 187-202.

- Gerard François-Marie (2013). « L'évaluation au service de la régulation des apprentissages : enjeux, nécessités et difficultés ». *Revue française de linguistique appliquée*, vol. XVIII, n° 1, p. 75-92.
- Hadji Charles (2012). *Faut-il avoir peur de l'évaluation ?* Bruxelles : De Boeck.
- Harlen Wynne (2006). « On the relationship between assessment for formative and summative purposes ». In Gardner John (dir.), *Assessment and Learning*. Thousand Oaks : Sage Publications, p. 87-102.
- Hattie John & Timperley Helen (2007). « The power of feedback ». *Review of Educational Research*, vol. 77, n° 1, p. 81-112.
- Hazard Brigitte, Gauthier Roger-François & Nicodeme Raymond (dir.) (2012). *La mise en œuvre du livret personnel de compétences dans les collèges*. Rapport conjoint IGEN/IGAENR. Paris : Ministère de l'Éducation nationale.
- Herman Joan L., Osmundson Ellen & Silver David (2010). *Capturing quality in formative assessment practice: Measurement challenges* (CRESST Report 770). Los Angeles : National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Houchot Alain, Thollon Frédéric, Charbonnier Daniel et al. (2013). *La notation et l'évaluation des élèves éclairées par des comparaisons internationales*. Paris : Inspection générale de l'éducation nationale.
- Kerrero Christophe (2014). Conférence introductive au colloque « *Évaluation des élèves et démocratie* », SNPDEN, ENSMM, Besançon, 14 mai.
- Lafortune Louise & Allal Linda (dir.) (2008). *Jugement professionnel en évaluation : Pratiques enseignants au Québec et à Genève*. Sainte-Foy : Presses de l'université du Québec.
- Lassen Steen (2013). « Evaluation and assessment of key competences in Denmark ». *Conference on better assessment and evaluation to improve teaching and learning*, 20 mars 2013, Dublin.
- Looney Janet W. (2011). « Integrating Formative and Summative Assessment: Progress Toward a Seamless System ? ». *OECD Education Working Papers*, n° 58.
- Mangez Éric (2008). *Réformer les contenus d'enseignement*. Paris : Presses universitaires de France.
- Marcoux Géry, Fagnant Annick, Loyer Nathalie & Ndinga Pascal (2014). « L'évaluation diagnostique des compétences à l'école obligatoire ». In Dierendonck Christophe, Loarer Even, Rey Bernard (dir.). *L'évaluation des compétences en milieu scolaire et en milieu professionnel*. Bruxelles: De Boeck, p. 117-125.
- Maulini Olivier (2003). « L'école de la mesure. Rangs, notes et classements dans l'histoire de l'enseignement ». *L'éducateur*, mars, p. 33-37.
- Maulini Olivier (2012). « Resserrer ou justifier les classements ? L'évaluation scolaire entre deux injonctions ». *Diversité*, n° 169, p. 133-137.
- Merle Pierre (1998). *Sociologie de l'évaluation scolaire*. Paris : Presses universitaires de France.
- Merle Pierre (2005). *L'élève humilié : L'école, un espace de non-droit ?* Paris : Presses universitaires de France.
- Merle Pierre (2007). *Les notes. Secrets de fabrication*. Paris : Presses universitaires de France.
- Merle Pierre (2012a). « L'évaluation par les notes : quelle fiabilité et quelles réformes ? ». *Regards croisés sur l'économie*, vol. 2, n° 12, p. 218-230.
- Merle Pierre (2012b). « L'évaluation des élèves. Une modélisation interactionniste des pratiques professorales ». In Lucie Mottier Lopez & Gérard Figari (dir.). *Modélisations de l'évaluation en éducation*. Bruxelles : De Boeck, p. 27-43.
- Millon-Fauré Karine (2013). « Processus de négociation didactique et mesure du niveau des élèves : des fonctions concurrentes de l'évaluation ». *Carrefours de l'éducation*, n° 36, p. 149-166.
- Ministère de l'Éducation de l'Ontario (2009). *Évaluer, différencier... Réussir !* Toronto : Centre franco-ontarien de ressources pédagogiques.



- Ministère de l'Éducation de l'Ontario (2010). *Faire croître le succès : Évaluation et communication du rendement des élèves fréquentant les écoles de l'Ontario*. Toronto : Le Ministère.
- Moisan Jacques (2011). « Enseignement par compétences : former et évaluer ». In Pré-actes du *Colloque du MCLCM*, Paris, 14 mai, p. 13-14.
- Mons Nathalie (2009). « Effets théoriques et réels des politiques d'évaluation standardisée ». *Revue française de pédagogie*, n° 169, p. 99-140.
- Mottier Lopez Lucie (2013). « De la mesure à l'évaluation collaborative en éducation ». *Revue française d'administration publique*, n° 148, p. 939-952.
- Mottier Lopez Lucie & Allal Linda (2010). « Le jugement professionnel en évaluation : quelles triangulations méthodologiques et théoriques ? ». In Paquay Léopold, Van Nieuwenhoven Catherine & Wouters Pascale (dir.). *L'évaluation, levier du développement professionnel ? Tensions, dispositifs, perspectives*. Bruxelles : De Boeck, p. 237-249.
- Mottier Lopez Lucie, Tessaro Walther, Dechamboux Lionel & Villabona Fernando Morales (2012). « La modération sociale : un dispositif soutenant l'émergence de savoirs négociés sur l'évaluation certificative des apprentissages des élèves ». *Questions vives*, vol. 6, n° 18, septembre, p. 1-15.
- Muller François & Normand Romuald (2013). *École : la grande transformation ? : les clés de la réussite*. Issy-les-Moulineaux : ESF.
- Nusche Deborah, Radinger Thomas, Santiago Paulo & Shewbridge Claire (2013). « Synergies for better learning. An international perspective on evaluation and assessment ». *OECD Reviews of Evaluation and Assessment in Education*. Paris : OCDE.
- Oswald, Steven G. (2013). *Identifying formative assessment in classroom instruction: Creating an instrument to observe use of formative assessment in practice. Dissertation in education in curriculum and instruction*. Boise : Boise State University.
- Pepper David (2013). *KeyCoNet 2013 literature review: Assessment for key competences*. Bruxelles : Key competence network on school education.
- Petkova Iliana (2013). « Les conceptions de l'évaluation des enseignants en Bulgarie ». In *Évaluation et autoévaluation : quels espaces de formation ?* Colloque de l'ADMEE, Fribourg, 9-11 janvier.
- Rémond Martine (2008). « Qu'est-ce qu'évaluer ? ». Communication présentée au séminaire « Accompagner les enseignants dans la mise en œuvre du socle commun », IUFM de Livry, 14 mai. Lyon : INRP-ENS de Lyon.
- Resnik Julia (2013). « L'ère de l'évaluation aux États-Unis et l'intégration du baccalauréat international comme dispositif pour améliorer les lycées publics ». *Carrefours de l'éducation*, n° 36, p. 77-94.
- Reuter Yves (2013). *Panser l'erreur à l'école*. Villeneuve d'Ascq : Presses universitaires du Septentrion.
- Reverdy Catherine (2013). « Des projets pour mieux apprendre ? » *Dossier de veille de l'IFÉ*, n° 82, février. Lyon : ENS de Lyon.
- Rey Bernard (2012). « Culture de l'évaluation et exigences éthiques ». *Education, Sciences & Society*, vol. 2, n° 2, p. 97-108.
- Rey Bernard (2014). « Compétence et évaluation en milieu scolaire : une relation complexe ». In Dierendonck Christophe, Loarer Even, Rey Bernard (dir.). *L'évaluation des compétences en milieu scolaire et en milieu professionnel*. Bruxelles : De Boeck, p. 25-34.
- Rey Bernard (2012). « Culture de l'évaluation et exigences éthiques ». *Education Sciences & Society*, vol. 2, n° 2, p. 97-108.
- Rey Olivier (2011). « PISA : ce que l'on en sait et ce que l'on en fait ». *Dossier de veille de l'IFÉ*, n° 66, octobre. Lyon : ENS de Lyon.
- Rey Olivier (2012). « Le défi de l'évaluation des compétences ». *Dossier de veille de l'IFÉ*, n° 76, juin. Lyon : ENS de Lyon.



- Roegiers Xavier (2014). « Quelle évaluation des compétences, au service de quel projet pour l'école ? » In Dierendonck Christophe, Loarer Even, Rey Bernard (dir.). *L'évaluation des compétences en milieu scolaire et en milieu professionnel*. Bruxelles: De Boeck, p. 71-84.
- Ruiz-Primo Maria Araceli (2011). « Informal formative assessment : The role of instructional dialogues in assessing students' learning ». *Studies in Educational Evaluation*, n° 37, p. 15-24.
- Rupp André A. & Lesaux Nonie K. (2006). « Meeting expectations? An empirical investigation of a standards-based assessment of reading comprehension ». *Educational Evaluation and Policy Analysis*, vol. 28, n° 4, p. 315-333.
- Scallon Gérard (2007). *L'évaluation des apprentissages dans une approche par compétences* [2<sup>e</sup> éd.]. Bruxelles : De Boeck.
- Serres Chantal (2014). « Avant-propos ». In Dierendonck Christophe, Loarer Even, Rey Bernard (dir.). *L'évaluation des compétences en milieu scolaire et en milieu professionnel*. Bruxelles: De Boeck, p. 17-22.
- Shavelson Richard, Young Donald, Ayala, Carlos et al. (2008). « On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers ». *Applied measurement in education*, vol. 21, n° 4, p. 295-314.
- Shepard Lorrie (2005). « Formative assessment: Caveat emptor ». In *ETS Invitational Conference 2005 : The future of assessment: Shaping teaching and learning*, New-York, octobre.
- Stobart Gordon (2011). « L'évaluation pour les apprentissages : d'une expérimentation locale à une politique nationale ». *Revue française de pédagogie*, n° 174, p. 41-48.
- Suchaut Bruno (2008). « La loterie des notes au bac. Un réexamen de l'arbitraire de la notation des élèves ». *Les documents de travail de l'IRE-DU*, n° 2008-3.
- Tessaro Walther (2013). « Améliorer la qualité des pratiques évaluatives des enseignants : une articulation entre formation initiale et formation continue ». *Bulletin de la Haute école pédagogique de Bern, du Jura et de Neuchâtel*, n° 21, février, p. 8-9.
- Timmis Sue, Oldfield Alison, Broadfoot Patricia & Sutherland Rosamund (2012). « Where is the cutting edge of research in e-Assessment? Exploring the landscape and potential for wider transformation ». In Whitelock, D., Warburton, W., Wills, G., and Gilbert, L. (dir.). *CAA 2012 International Conference*, University of Southampton.
- Trouilloud David & Sarrazin Philippe (2011). « Les attentes des enseignants : une menace pour l'évaluation ? ». In Butera Fabrizio, Buchs Céline, Darnon Cécile (dir.). *L'évaluation, une menace ?* Paris : Presses universitaires de France, p. 67-75.
- Vial Michel (2012). *Se repérer dans les modèles de l'évaluation : méthodes, dispositifs, outils*. Bruxelles : De Boeck.









► **Pour citer ce dossier :**

Rey Olivier, Feyfant Annie (2014). *Évaluer pour (mieux) former. Dossier de veille de l'IFÉ*, n°94, septembre. Lyon : ENS de Lyon  
En ligne : <http://ife.ens-lyon.fr/vst/DA/detailsDossier.php?parent=accueil&dossier=94&lang=fr>

► **Retrouvez les derniers Dossiers de veille de l'IFÉ :**

- Endrizzi Laure (2014). *La qualité de l'enseignement : un engagement des établissements, avec les étudiants ?* Dossier de veille de l'IFÉ, n° 93, juin. Lyon : ENS de Lyon.  
En ligne : <http://ife.ens-lyon.fr/vst/DA/detailsDossier.php?parent=accueil&dossier=93&lang=fr>
- Gausse Marie (2014). *Petite enfance : de l'éducation à la scolarisation.* Dossier de veille de l'IFÉ, n° 92, avril. Lyon : ENS de Lyon.  
En ligne : <http://ife.ens-lyon.fr/vst/DA/detailsDossier.php?parent=accueil&dossier=92&lang=fr>
- Reverdy Catherine (2014). *De l'université à la vie active.* Dossier de veille de l'IFÉ, n° 91, mars. Lyon : ENS de Lyon.  
En ligne : <http://ife.ens-lyon.fr/vst/DA/detailsDossier.php?parent=accueil&dossier=91&lang=fr>

► **Abonnez-vous aux Dossiers de veille de l'IFÉ :**

<http://ife.ens-lyon.fr/vst/abonnement.php>

© École normale supérieure de Lyon  
Institut français de l'Éducation  
Veille et Analyses

15 parvis René-Descartes BP 7000 – 69342 Lyon cedex 07  
[veille.scientifique@ens-lyon.fr](mailto:veille.scientifique@ens-lyon.fr)  
Standard : +33 (04) 26 73 11 24  
ISSN 2272-0774