

L'évaluation au cœur des apprentissages

Laure Endrizzi, Olivier Rey

▶ **To cite this version:**

Laure Endrizzi, Olivier Rey. L'évaluation au cœur des apprentissages. 2008. halshs-00473757

HAL Id: halshs-00473757
https://shs.hal.science/halshs-00473757

Submitted on 19 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Service de Veille scientifique et technologique **http://www.inrp.fr/vst**

Dossier d'actualité
(ancien titre : Lettre d'information)
n° 39 – novembre 2008

vous abonner
consulter le dossier en ligne

L'évaluation au cœur des apprentissages

Par Laure Endrizzi et Olivier Rey

Le traitement scolaire des connaissances apparaît généralement inséparable des procédures de contrôle de ces connaissances ; les examens et les différentes formes d'évaluation tiennent donc une place centrale dans les dispositifs éducatifs, que ce soit pour attester d'une formation (évaluation certificative), vérifier si des savoirs et compétences sont acquis (évaluation sommative), apprécier le niveau atteint et les potentialités à poursuivre dans telle ou telle voie (évaluation pronostique), jauger du niveau atteint par une classe d'âge ou une population scolaire (évaluation diagnostique), etc.

Toutes ces évaluations sont importantes et parfois omniprésentes dans le fonctionnement des systèmes éducatifs. Elles induisent le problème de la place de la notation et celui de la façon dont elles pèsent sur le pilotage des programmes et du curriculum. Elles sont aussi profondément interrogées par le développement des nouvelles approches par compétences, qui remettent en cause les formes et les outils traditionnels de l'évaluation.

Dans le cadre de ce dossier, nous avons généralement privilégié la question de l'évaluation pour l'apprentissage, parfois aussi nommée évaluation « formative », plutôt que celle de l'évaluation des apprentissages. Autrement dit, comment l'évaluation intervient-elle dans le cours même du processus d'apprentissage, pour le faciliter, le réorienter ou le réguler, plutôt que comme une sanction *a posteriori*.

Dans ce cadre, une attention particulière sera accordée à l'engagement de l'élève dans les processus d'évaluation, y compris aux différentes dimensions que peut prendre l'auto-évaluation. La question du portfolio dans l'éducation, qui concentre de nombreuses attentes sur ces sujets, fera l'objet d'un développement spécifique.

D'une évaluation à l'autre | L'évaluation formative ou évaluation pour l'apprentissage | Le feedback des élèves pour réguler les enseignements ? | L'évaluation des compétences | Le portfolio : un outil réflexif au service de l'évaluation ? | Bibliographie.

Avertissements au lecteur :

- *la plupart des liens renvoient vers les fiches correspondantes de notre base bibliographique collaborative, qui comprennent 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) ;*
- *sauf indication contraire, toutes les traductions comprises dans ce* Dossier d'actualité *ont été réalisées par les rédacteurs ;*
- *vous pouvez faire part de vos réactions à ce* Dossier *en laissant un commentaire sous l'article correspondant dans notre blog : « Écrans de veille en éducation ».*

D'une évaluation à l'autre

De quelle évaluation parle-t-on ?

Peu de processus semblent 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.

Pourtant, si l'éducation existe sous une forme ou sous une autre dans la plupart des sociétés humaines depuis leurs débuts, souligne P. Broadfoot, ce n'est que plus récemment que l'évaluation est devenue pour sa part un domaine professionnel, central et sophistiqué dans le contexte de la scolarisation de masse.

Le développement d'une évaluation formalisée et systématique a évidemment à voir avec la légitimation des positions sociales par le diplôme, la promotion du mérite académique qui a marqué la construction des États modernes et l'utilisation des évaluations comme outil de contrôle institutionnel à tous les niveaux (Broadfoot, 2007).

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, une évaluation internationale des acquis des élèves comme PISA, 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.

À cette simple énumération, on saisit immédiatement que l'on ne parle pas vraiment de processus identiques, même s'il est question à chaque fois de mesure ou de jugement, et même s'il est évident que l'on perçoit le continuum qui peut exister entre l'évaluation « individuelle » de l'élève et celle plus « institutionnelle » qui intervient avec l'utilisation des tests standardisés aux États-Unis (Berlinet & Amrein, 2002) ou celle des évaluations internationales dans les systèmes éducatifs (Emin in Baillat, De Ketele, Paquay & Thélot, 2008 ; Cytermann & Demeuse, 2005).

On peut également traduire cela par 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'évaluations significativement différentes. Ainsi, il est de tradition de distinguer l'évaluation selon les fonctions qu'elle remplit dans l'apprentissage : diagnostique, pronostique, sommative, certificative, formative...

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).

On peut également catégoriser l'évaluation selon les objets qu'elle évalue (savoirs, habiletés, connaissances, compétences…) ou encore selon les outils mis en œuvre (contrôles écrits, grilles d'observation, portfolios). À l'heure où plusieurs pays déploient un agenda pour développer l'évaluation assistée par ordinateur (*computer-based assessment*), voire pour migrer tests et/ou examens sur support numérique (Scheuermann & Guimarães Pereira, 2008), les champs à investiguer en matière d'évaluation se démultiplient ; ce sont des questionnements techniques qui émergent (quels outils ?), mais aussi méthodologiques (comment ?), éthiques (jusqu'où ?) et psycho-cognitifs (quel impact sur les résultats ?). Dans ce contexte, l'articulation entre apprentissage, évaluation et TIC constitue un terrain d'exploration encore relativement vierge.

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 diplômant 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.

Si la certification a longtemps constitué le seul débouché de l'évaluation, la démocratisation scolaire qui s'est développée depuis les années 60 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, dimension privilégiée par notre dossier.

On le constate aussi selon Scallon dans le vocabulaire utilisé dans la recherche de langue anglaise, dans laquelle les termes de « *measure* » voire de « *testing* » ont quasiment disparu, et celui d'« *evaluation* » est devenu 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, selon G. Scallon, même si cela a constitué pendant une période l'horizon de la docimologie.

En outre, certains chercheurs, comme B. Rey, soulignent que quels que soient les raffinements techniques de l'évaluation, « *les dimensions que l'on mesure et que l'on prend en compte sont l'effet du choix d'un sujet* » (Rey in Baillat, De Ketele, Paquay & Thélot, 2008, p. 60).

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 l'abondant dossier en ligne réalisé par Jacques Nimier : « Cette évaluation impossible et pourtant nécessaire »).

La notation, horizon indépassable de l'évaluation ?

Les notes concentrent en effet 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.

Ainsi, 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 au Québec : 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 ainsi 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 90 a dû réintégrer la possibilité d'évaluations chiffrées en 2004, sous la pression là encore des familles.

Eric 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 formatives « qui ne comptent pas » (la question de l'évaluation formative étant réduite au débat sur le fait de noter ou de ne pas noter).

Certains chercheurs regrettent pourtant que, formés à la pédagogie par objectifs, les évaluateurs dans le système éducatif élaborent des grilles qui visent à limiter l'arbitraire de leurs décisions et rendre non contestables les notes attribuées. Ce faisant, ces grilles atomisent les différents éléments constitutifs des compétences, au risque d'en dénaturer totalement l'approche intégrative. D'autres pratiques de notation sont repérables mais ont du mal à s'implanter dans les classes, regrettent-ils (Dauvisis, in Hatano & Lemaître, 2007).

En France, le collège expérimental Clisthène (Bordeaux) a en revanche réussi à combiner l'utilisation des notes traditionnelles pour les savoirs à des évaluations de niveaux de compétences (symbolisés par les couleurs des feux), pour déterminer si un savoir-faire est acquis, en cours d'acquisition ou non-acquis (Cédelle, 2008). L'originalité de l'expérience est sans doute aussi la restitution individualisée à la famille d'un bulletin trimestriel particulièrement dense (4 pages), mêlant appréciations générales, notes « sur 20 » et appréciations détaillées des compétences elles-mêmes décomposées (avec toujours les repères visuels de couleur).

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

Pierre Merle, qui a mené une enquête minutieuse sur la « fabrication » des notes dans l'enseignement secondaire (Merle, 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.

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...

Enfin, il y a évidemment la fixation arbitraire de notes plancher et plafond, rarement liée à une logique de performance scolaire, mais souvent justifiée par des préférences personnelles : seuil psychologique, soutien de la motivation, maintien d'exigences élevées... Dans tous les cas, le refus des notes « extrêmes » est utilisé pour cadrer la relation et l'autorité pédagogique. Ce phénomène est proche de ce que le mathématicien André Antibi a appelé depuis 1983 « la constante macabre » et qui est désormais bien connu comme la répartition constante de bonnes et mauvaises notes de part et d'autre de la moyenne, quelle que soit la performance « réelle » des élèves.

Malgré cela, P. Merle propose, plus qu'un abandon de la notation, un effort continu d'amélioration des techniques pour assurer une plus grande justice scolaire car, selon lui, des épreuves anonymes telles que celles du Bac assurent une plus grande équité que le contrôle continu, plus sensible aux biais induits par le jugement scolaire professoral.

L'évaluation outil de pilotage du curriculum

Si, dans une vision naïve, l'évaluation se contente d'enregistrer ou de sanctionner ce qui a été appris, de nombreux travaux ont montré qu'en fait, les contenus et programmes d'enseignement sont largement influencés voire pilotés par leurs modes d'évaluation.

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 (coefficients différentiels notamment) ou à quel point il est difficile d'obtenir qu'un enseignement « sans évaluation » soit pris en considération. **Un travers souvent dénoncé de la réification de l'évaluation dans l'éducation, qui conduit à valoriser ce qui est mesurable plutôt qu'à mesurer ce qui a de la valeur** (Broadfoot, 2007, p. 155).

Au-delà de la polarisation de l'intérêt des élèves, l'évaluation guide plus globalement l'enseignement, y compris ses objectifs, ses stratégies et même les tâches utilisées pour le développement de diverses habiletés, qui sont dérivées de l'anticipation de l'évaluation. À partir d'une enquête sur l'enseignement de l'anglais langue seconde en Inde, par exemple, un chercheur montre que la nature de l'évaluation aux niveaux les plus élevés du cycle de formation est cruciale dans la formation des perceptions des enseignants quant à ce qu'il est important d'enseigner et d'évaluer, et quant aux habiletés et compétences qui doivent être développées et encouragées (Agrawal, 2004).

Dès lors, les effets induits peuvent être lourds de conséquences, quand un système met au centre les résultats des élèves à des tests d'évaluation normés, comme c'est le cas dans certains pays anglo-saxons depuis la fin du XXe siècle. Loin d'apporter les progrès attendus en matière d'apprentissages, cette centration sur les tests produit, en revanche, de la tension et surtout une déformation de l'enseignement réduit à une préparation intensive aux tests, le fameux « teaching to the test », comme le montre l'étude comparée des processus liés aux tests dans 18 états américains (Berliner & Amrein, 2002). Dans son intervention au Congrès de l'association américaine des chercheurs en éducation de 2007, sa présidente, Eva L. Baker, s'interrogeait ainsi sur la nécessité de mettre en place un système de « qualifications » qui prenne mieux en compte le développement personnel et la réalité des apprentissages, à côté des objectifs de performance poursuivis par les tests actuels (Baker, 2007). E. Baker était alors aussi directrice de l'important Centre national de recherche sur l'évaluation, les standards et les tests de l'Université de Californie (CRESST).

La montée en puissance, depuis les années 80, 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*).

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 menées conjointement de 1989 à 1997 (Broadfoot & Pollard in Lauder, Brown, Dillabourgh & Halsey, 2006). Selon eux, via l'imposition 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.

En tout état de cause, les dispositifs d'évaluation jouent un rôle central pour changer les règles qui régissent les idées et les pratiques éducatives, en rendant de plus en plus difficile l'expression même d'une conception alternative des objectifs éducatifs souhaitables, en concluent Pollard et Broadfoot.

On peut aussi trouver des dispositifs d'évaluation qui oscillent entre une évaluation utilisée comme un soutien à l'apprentissage et au développement personnel et une évaluation orientée vers les indicateurs de performance, comme le dispositif SIMCE (*Sistema nacional de medición de resultados de aprendizaje del Ministerio de Educación*) au Chili (Cariola & Meckes, 2008). Le SIMCE, créé sous le gouvernement militaire, visait à proposer un outil de mesure de la qualité des établissements, pour améliorer la concurrence scolaire dans l'optique de développer un marché éducatif. Durant les premières années du gouvernement démocratique, de 1990 à 1995, le système a été utilisé au contraire pour décrire les inégalités entre écoles afin de guider les politiques publiques compensatoires et d'aider les enseignants à améliorer leurs résultats. Pendant la seconde moitié des années 90 le système a été davantage utilisé comme outil de contrôle de la politique éducative, avec une publication des résultats des établissements et des conséquences sur les primes des enseignants et les dotations des écoles. Depuis le début du XXIe siècle, le SIMCE s'oriente dans une optique plus marquée de reddition des comptes (*accountability*) en termes d'obtention de résultats d'apprentissages nationaux standardisés, laissant de côté les indicateurs initiaux de développement personnel et d'intégration scolaire des élèves, au profit de la seule performance mesurée par les résultats.

L'évaluation formative ou évaluation pour l'apprentissage

De la rétroaction à la régulation interactive

L'évaluation formative concentre une grande partie des réflexions actuelles sur les évaluations réalisées pendant l'apprentissage et *pour* l'apprentissage (par contraste avec les évaluations *des* apprentissages), comme en témoigne par exemple la parution de l'ouvrage du CERI-OCDE sur la question (CERI, 2005).

La notion d'évaluation formative a été introduite par Michael Scriven en 1967 puis popularisée par Bloom en 1971, quand il l'a intégrée à son modèle de la « pédagogie de maîtrise » (Allal & Mottier-Lopez, in CERI, 2005). À l'époque, évaluation formative et évaluation sommative se différencient essentiellement par leurs objectifs ou les utilisations qu'on peut en faire : à l'évaluation sommative qui indique si certaines connaissances ou apprentissages ont été acquis par l'élève, l'évaluation formative ajoute un « feedback » ou rétroaction qui consiste à fournir des informations sur le degré d'acquisition et les erreurs commises, afin de reprendre, d'approfondir ou de corriger l'apprentissage en conséquence.

Certains chercheurs estiment que cette théorie originelle est toujours valide et qu'il serait peu pertinent d'autonomiser artificiellement l'évaluation formative voire de créer une fausse opposition entre évaluation « sommative » et « formative » (Taras, 2005). D'autres préfèrent souligner les modifications du concept depuis une vingtaine d'années, car si l'évaluation formative était initialement intégrée dans un processus de type linéaire « apprentissage-évaluation-correctif », elle s'est progressivement insérée dans une conception de la régulation interactive des apprentissages, dans laquelle le guidage de la situation d'enseignement s'ajuste en continu aux retours des élèves (Scallon, in Grégoire, 2008).

Cette idée de régulation a été développée, selon Linda Allal, pour désigner spécifiquement les mécanismes qui assurent le guidage, le contrôle et l'ajustement des activités cognitives, affectives et sociales, ainsi que leur articulation (Allal, in Allal & Mottier-Lopez, 2007). La psychologie cognitive, également, s'est intéressée à la démarche, pour comprendre les opérations mentales à la source des réussites ou des échecs, identifier les « préconceptions » problématiques, et donc aller plus loin que les analyses traditionnelles de type comportementalistes (Grégoire, 2008).

À partir d'une théorisation socioculturelle, des chercheurs s'attachent aussi à mieux distinguer les nombreuses pratiques associées à l'évaluation formative, dans un objectif avoué de dépasser les effets d'étiquetage superficiels et d'aller jusqu'au bout de la démarche visant à faire de l'élève le sujet plus que l'objet de l'éducation (Pryor & Crossouard, 2008).

Une enquête d'Eleanor Hargreaves (2005) avait en effet montré que les enseignants anglais ont des conceptions différentes de l'apprentissage, derrière une commune adhésion déclarée à l'évaluation formative. Pour les uns, il s'agissait en fait d'une évaluation très comparable à une mesure traditionnelle, pendant que pour les autres, il s'agissait d'un processus de co-investigation. Comme par ailleurs la multiplication des tests standardisés ne favorise pas une conception très complexe de l'évaluation formative, Pryor et Crossouard ont cherché, sur la base de recherches expérimentées avec les praticiens, à dégager les pratiques envisageables et constitutives d'une réelle évaluation formative. Cela les a amenés à regrouper un ensemble de pratiques autour de deux évaluations formatives idéal-typiques. D'une part le modèle convergent, d'inspiration largement comportementaliste (stimulus-réponse), dans lequel le feedback vise essentiellement à donner le signal de la bonne réponse à des élèves susceptibles de savoir utiliser ces indications grâce à leurs bases culturelles. D'autre part le modèle divergent, où le feedback exploratoire, informationnel voire provocateur a pour objectif d'engager l'élève au-delà d'une simple correction d'erreur, pour aider l'enseignant comme l'élève à mieux identifier quels sont les savoirs en jeu, les malentendus et les pré-requis liés, afin de construire l'apprentissage comme une collaboration.

C'est cette deuxième acception qui est au cœur des travaux de Black et Wiliam, depuis leur célèbre revue de littérature (1998) jusqu'aux recherches expérimentales qu'ils poursuivent depuis vingt ans. Le projet KMOFAP (*King's-Medway-Oxfordshire Formative Assessment Project*) en particulier représente un jalon important dans la mise en évidence de l'impact de l'évaluation formative sur les performances des élèves et consécutivement dans la dissémination des pratiques. À l'issue de cette expérience, quatre types de pratiques perfectionnées ont été identifiées (Black & Wiliam in CERI, 2005) :

- améliorer la dimension interactive du feedback oral pour que le dialogue avec la classe soit plus riche et plus utile. Ceci passe notamment par l'allongement du délai de réponse accordé aux élèves, un traitement sérieux des réponses de ces derniers pour mieux identifier leurs lacunes ou idées fausses, une formulation des questions et des réponses des enseignants plus centrées sur les besoins des élèves et l'amélioration de leur compréhension ;
- mettre l'accent sur les commentaires, dans le feedback écrit, jusqu'à supprimer la notation dans certains cas, pour que la note ne subordonne pas les commentaires, ce qui est usuellement constaté ;
- utiliser l'évaluation entre pairs et l'auto-évaluation pour que les élèves comprennent mieux les objectifs à atteindre derrière les tâches à réaliser, et les moyens pour atteindre ces objectifs, en s'appropriant les critères d'évaluation de la réussite ;
- exploiter les tests d'évaluation sommative à des fins formatives, en demandant aux élèves d'assortir de feux de signalisation les sujets sur lesquels vont porter les contrôles, afin de mettre en place de vraies stratégies de préparation ou encore d'élaborer eux-mêmes des questions d'examens pour acquérir une meilleure vision du sujet.

Ces premiers résultats ont contribué à la popularité du concept dans les pays anglo-saxons, et ont notamment conduit à l'adoption en 2004 de l'évaluation pour l'apprentissage (*assessment for learning*) comme l'un des domaines clés du programme anglais pour les élèves du Key Stage 3 (11-14 ans) et au lancement du programme d'accompagnement Assessment is for Learning (AifL) en Ecosse, également en 2004. En Angleterre aujourd'hui, l'évaluation pour l'apprentissage fait partie intégrante du programme de personnalisation des apprentissages (Personalised Learning) en cours de généralisation dans l'enseignement obligatoire.

Entre auto-évaluation et co-évaluation : quelles synergies ?

L'influence de l'évaluation sur l'engagement des élèves dans l'apprentissage n'est plus à questionner. Dès la fin des années 80, les travaux de Crooks (1988) ont bien montré comment l'évaluation « *guide leur jugement sur ce qu'il est important d'apprendre, affecte leur motivation et leurs représentations en matière de compétence, structure leur organisation en termes de travail personnel, renforce leurs apprentissages et impacte durablement le développement d'aptitudes et de stratégies d'apprentissage* ».

Avec la croissance exponentielle des opportunités d'apprendre dans le monde numérique d'aujourd'hui, le besoin de maîtrise de l'apprentissage et de l'évaluation requiert une approche « durable » : pour savoir gérer son rapport au savoir dans un contexte scolaire formel et dans un environnement social plus informel, autrement dit pour savoir tout au long de la vie évaluer ses forces et ses faiblesses avant de décider de la prochaine étape.

S'inscrire dans cette perspective, c'est reconnaître avec P. Broadfoot (2007 p. 135-136) que « ***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* ».

En tout état de cause, cette 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). Dans les études de cas présentées dans l'ouvrage, cette dimension participative peut prendre plusieurs formes, utiliser différentes techniques et intervenir à différents moments pédagogiques. Certaines pratiques régulières peuvent toutefois être observées :

- 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).

Au-delà du simple niveau descriptif, la manière dont ces pratiques de classe influent à la fois sur l'expérience scolaire et sur les résultats de l'apprentissage conserve une part d'opacité. En quoi l'implication des élèves dans l'évaluation, et en particulier le fait d'encourager l'auto-évaluation et de l'évaluation entre pairs, est-elle bénéfique et dans quelles conditions ?

Les recherches de Black et Wiliam sur l'auto-évaluation (in CERI, 2005) s'enracinent dans les travaux de Sadler (1989), montrant que la compréhension des objectifs d'apprentissage, et consécutivement l'appréciation des efforts à fournir pour les atteindre, sont au cœur de l'apprentissage. Dans cette démarche auto-évaluative, une certaine **transparence des critères de l'évaluation** est donc nécessaire : c'est en parvenant à se construire une vision globale de la tâche à réaliser et des objectifs qui la sous-tendent que les élèves peuvent développer des compétences métacognitives suffisantes pour mieux gérer et maîtriser cette même tâche (White & Frederiksen, 1998).

Mais la construction de cette vue d'ensemble ne va pas de soi. Weeden et Winter ont bien montré, dans le cadre du projet LEARN, que les représentations des élèves sur l'évaluation étaient caractérisées d'une part par une forte dépendance à l'égard des pratiques des enseignants et d'autre part par un manque de perspectives : s'ils comprennent généralement ce que l'on attend d'eux à l'occasion d'une tâche particulière, ils ne parviennent pas à intégrer cette tâche particulière dans une vue plus globale (big picture), autrement dit **ils ne savent pas quel chemin ils ont déjà parcouru et combien il reste à parcourir pour atteindre leur destination.** (Weeden & Winter, 1999).

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, p. 69). 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. En outre, il semble que les élèves s'approprient plus facilement les critères d'évaluation en examinant un travail qui n'est pas le leur.

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 répandues (**formellement**) n'est d'ailleurs pas non plus attesté. Examinant le tutorat entre pairs, A. 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 partagé. L'idéal résiderait finalement dans un tutorat qui se fait « sans se dire ».

La récente revue de littérature réalisée par Sebba *et al.* (Sebba, 2008) pour le compte de l'EPPI-Centre de l'institut d'éducation de Londres confirme ce manque de recherches empiriques sur la co-évaluation, du simple fait que les 26 études retenues dans l'analyse sont majoritairement antérieures à 2000. Les bénéfices, associés indistinctement à l'auto-évaluation et la co-évaluation, sont essentiellement perceptibles sur **l'engagement des élèves dans l'apprentissage et le développement de l'estime de soi**, avec, dans une moindre mesure, une certaine influence sur les résultats scolaires. Ce type d'approche est favorisé par une culture de classe qui encourage le dialogue enseignant-élèves et par **l'évolution d'une relation de dépendance à une relation d'interdépendance** : l'enseignant ajuste alors sa pédagogie en fonction du feedback des élèves (et vice versa). Le contrôle de l'élève sur le processus d'évaluation ne semble pas clairement porteur de bénéfice, bien que la participation des élèves à l'élaboration des critères d'évaluation soit considérée comme importante.

Pour Nicol et MacFarlane-Dick (2006), la distinction entre auto-évaluation et co-évaluation est artificielle. L'une et l'autre sont au final constitutives de l'auto-régulation au cœur des apprentissages, et la question à privilégier est davantage celle du **feedback (interne et externe)** dans l'évaluation formative. Selon eux, le **modèle transmissif** du feedback, largement répandu, est à remettre en question. D'une part parce que la responsabilité de l'enseignant ne peut être seule engagée dans un processus où les élèves sont supposés prendre conscience de leurs apprentissages et qu'en tout état de cause la charge de travail induite pour l'enseignant dans ce cas n'est pas supportable ; d'autre part parce que le fait de transmettre un feedback ne peut en soi suffire à produire une action régulative, si les dimensions psycho-sociales ne sont pas prises en compte d'une manière ou d'une autre. L'activité de l'enseignant dans cette vision transmissive commune doit donc être minorée, au profit du rôle actif et central que les élèves ont à jouer, et concentrée sur les sept principes suivants, selon lesquels l'enseignant qui développe des bonnes pratiques en matière de feedback :

1. aide à comprendre ce que signifie une bonne performance (buts, critères, standards) ;

2. facilite le développement de l'auto-évaluation et de la réflexivité ;
3. adresse des informations de qualité aux élèves sur leurs apprentissages ;
4. encourage les élèves à dialoguer avec lui ou entre eux sur l'apprentissage ;
5. encourage la motivation et l'estime de soi des élèves ;
6. fournit des opportunités pour réduire la distance entre les performances actuelles et les performances désirées ;
7. fournit des informations à ses collègues enseignants, qui peuvent être exploitées pour modeler leur enseignement.

De l'auto-régulation au savoir apprendre

Le programme Assessement is for Learning (AifL) du ministère de l'éducation écossais distingue trois types de relations entre évaluation et apprentissage (*assessment for learning*, *assessment as learning*, *assessment of learning*), positionnant ainsi le savoir-apprendre comme une des composantes clés du programme. Implicite dans les principes fondamentaux établis par l'OCDE, ce focus sur le savoir-apprendre figure également au nombre des 10 principes clés de l'évaluation formative définis par l'Assessment Reform Group (ARG) en 2002 : « *assessment for learning should focus on how students learn* ».

Si les fondations mêmes de l'évaluation formative portent sur ce « pari » que les élèves doivent être **capables d'évaluer et de réviser leur propre travail, et donc d'identifier les prochaines étapes de l'apprentissage**, ce qu'il faut entendre concrètement par « savoir-apprendre » n'est pas évident. Faire participer les élèves au diagnostic de leur propre style d'apprentissage, par exemple à l'aide de la théorie des intelligences multiples d'Howard Gardner, peut soutenir la réalisation de cet objectif ; utiliser la technique des feux tricolores pour permettre à chaque élève de diagnostiquer ses forces et ses faiblesses participe également à cette sensibilisation au savoir-apprendre (CERI, 2005). Mais la promotion de simples techniques offre des perspectives limitées. Et l'auto-régulation ne va pas de soi, comme l'ont montré les travaux de Zimmerman et Schunk, notamment cités par Nicol et MacFarlane (2006).

Comment ce savoir-apprendre peut-il s'enseigner, se mesurer ? Quels défis soulève l'acquisition d'une telle compétence à l'école et quelles relations peut-on établir entre savoir-apprendre et évaluation pour l'apprentissage ? C'est à l'ensemble de ces questions, encore largement inexplorées, que les chercheurs du projet Learning how to learn – in classrooms, schools and networks (LHTL) ont tenté de répondre, en se fixant pour objectif la réalisation d'un modèle pédagogique du savoir-apprendre exploitable directement par les enseignants et les élèves.

Dans le numéro spécial de la revue *Research Papers in Education* consacré au projet (2006), l'article de D. Pedder (*Organizational conditions that foster successful classroom promotion of learning how to learn*) montre que l'introduction de cette dimension du savoir-apprendre dans les classes est guidée inégalement par trois objectifs : rendre l'apprentissage explicite, promouvoir l'autonomie des élèves et veiller aux progrès des résultats scolaires. L'accompagnement fourni par l'équipe du projet LHTL permet une diminution progressive du poids des performances, au profit d'un focus plus évident sur la **promotion de l'autonomie**, rendant les arbitrages des enseignants moins contraints par les programmes scolaires.

La méthodologie n'est toutefois pas généralisable, comme l'ont montré P. Black *et al.* dans un autre article du même numéro (*School pupils' beliefs about learning*). La conception d'un instrument permettant de mesurer l'acquisition d'une compétence métacognitive n'a en effet pas abouti : les biais inhérents à l'analyse des discours des élèves (réponses qui se contredisent mutuellement ou qui sont jugées trop conformes à ce qui serait attendu, etc.), les difficultés rencontrées pour traduire de façon abstraite leurs pratiques ou bien pour analyser leur propre engagement en liaison avec leur environnement scolaire, la complexité des paramètres à considérer pour concevoir un outil de mesure générique fiable, rendent en effet sa généralisation dans les écoles fortement improbable.

Une des principales conclusions de ce projet, qui fait directement écho à celles du projet contemporain Sustainable thinking classrooms, réside donc dans la nécessaire contextualisation de la compétence métacognitive. Autrement dit, développer une approche pédagogique du savoir-apprendre passerait par une inscription dans les disciplines : **enseigner (et mesurer) le savoir-apprendre, c'est enseigner (et mesurer) le savoir apprendre quelque chose**.

Le feedback des élèves pour réguler les enseignements ?

Via des consultations externes ?

Une certaine mobilisation existe dans les pays anglo-saxons en faveur de l'élève évaluateur, qui participe non seulement à sa propre évaluation, mais aussi à celle des enseignants et des services qui composent l'offre scolaire. Les digests du ministère de l'éducation sur le thème pupil voice en Angleterre, le programme SoundOut aux Etats Unis, le bimensuel des bonnes pratiques australiennes Connect ne sont que quelques exemples représentatifs de ce mouvement, fréquemment publicisé par les expressions « *pupil voice* » ou « *student voice* ».

Mais cette question du feedback des élèves n'est pas née avec les préoccupations démocratiques récentes liées au droit de l'enfant, revendiquées dans les mouvements de type « *student voice* ». Avec la démocratisation des

études supérieures dans les années 1980 et l'essor de la concurrence entre les établissements, émergent des interrogations sur l'adoption d'une démarche qualité, et corrélativement sur les méthodes à employer pour mesurer la satisfaction des étudiants (Harvey, 2001).

Aujourd'hui des enquêtes nationales sont administrées pour recueillir les points de vue des étudiants sur l'ensemble de leur expérience universitaire : la National survey of student engagement (NSSE) aux Etats Unis, le Course Experience Questionnaire (CEQ) en Australie et le programme expérimental anglais Student Listening Programme (SLP) par exemple ; d'autres dispositifs, tels que la National Student Survey (NSS) au Royaume Uni ou bien encore le Student Course Experience Questionnaire australien (SCEQ), sont davantage centrés sur l'expérience pédagogique, en particulier dans le premier cycle universitaire.

Les fondements d'une telle démarche n'ont cependant rien d'évident. La manière dont l'information est collectée et la temporalité (semestrielle, annuelle, bi-annuelle, etc.) dans laquelle s'inscrit cette collecte orientent nécessairement l'exploitation ultérieure des données. Et la manière dont cette exploitation est rendue publique sert là aussi de façon plus ou moins explicite différents enjeux. (Brennan & Williams, 2004 ; Gordon, 2005). Devant ces tensions entre démarche qualité, libre concurrence, obligation de résultats et révision des curriculums, la transition du « *self-assessment* » (l'élève s'auto-évalue) au « *self-evaluation* » (l'école s'auto-évalue), peut-elle contribuer à infléchir plus ou moins durablement les pratiques pédagogiques ? Autrement dit, peut-on réellement identifier l'influence de ces consultations sur les pratiques de classe d'une part et sur les attitudes et performances des élèves d'autre part ? Et s'il ne suffit pas aux élèves de donner leurs points de vue et aux enseignants d'en prendre connaissance pour que les bénéfices soient tangibles, dans quelles conditions cette approche s'avère-t-elle efficace ?

Un des premiers doutes qui semble pouvoir être levé porte sur la fiabilité et la pertinence du feedback des étudiants, comme le montre la revue de littérature de Prebble *et al.*, consacrée à l'évaluation des services d'appui offerts aux étudiants de premier cycle. Investiguant l'impact du feedback des étudiants sur la qualité globale de leurs enseignements, les auteurs parviennent à la conclusion suivante : « *les évaluations faites par les élèves comptent parmi les indicateurs les plus fiables et les plus accessibles de l'efficacité de l'enseignement. Bien utilisés, ils peuvent conduire à des améliorations qualitatives dans des pratiques d'enseignement et d'apprentissage* » (Prebble, 2005, p. 41). Ainsi, si la pertinence globale des évaluations étudiantes n'est pas à remettre en cause, il importe qu'elles ne constituent qu'un indicateur parmi d'autres et que la manière dont les données recueillies sont exploitées et restituées aux enseignants fasse l'objet d'une attention particulière, de façon à éviter tout malentendu. En effet, si nombre de recherches permettent aujourd'hui de comprendre les avantages et inconvénients associés aux différentes méthodes de collecte (Richardson, 2005 ; Hoban & Hastings, 2006), les travaux scientifiques concernant l'output, c'est-à-dire les modalités de restitution des résultats devant (ou non) stimuler l'action régulative, sont nettement moins précis.

De même si certaines expériences offrent à penser sur les meilleures méthodes pour « boucler » la rétroaction (*closing the loop*), les conditions pour une généralisation à l'ensemble des établissements d'enseignement supérieur, ne sont pas (encore) réunies (Brennan & Williams, 2004 ; Gordon, 2005). Là encore la question de l'adéquation entre les objectifs et les moyens est cruciale, car la démarche n'est nécessairement pas la même si la perspective est sommative ou bien si elle est formative ; autrement dit si l'objectif est de dresser un état des lieux à un moment donné du degré de satisfaction des élèves pour évaluer la qualité globale de l'offre pédagogique, ou bien s'il s'agit précisément de promouvoir une approche régulative requérant dès lors une intégration plus forte à la pédagogie (Gordon, 2005). Et l'une et l'autre des options n'ont nécessairement pas le même impact sur les pratiques pédagogiques, qu'il s'agisse de l'enseignement supérieur ou de l'enseignement scolaire.

Dans le quotidien des apprentissages ?

En Angleterre, les travaux fondateurs de Jean Rudduck dans le cadre du projet Consulting pupils about teaching and learning (2001-2003), ont mis en évidence les bénéfices croisés du feedback des élèves, si tant est qu'il ne s'agisse pas d'une démarche purement démagogique. Les résultats rendent compte d'un engagement renforcé des élèves dans la communauté éducative, d'une plus grande confiance en eux et d'une meilleure appréhension de leur identité d'apprenants ; les enseignants pour leur part voient leurs représentations évoluer et se sentent stimulés dans le renouvellement de leurs pratiques. Dans l'ouvrage publié avec J. Flutter (2004), les auteurs considèrent la participation de l'élève comme la pierre angulaire des dispositifs d'évaluation, ouvrant des perspectives de progression à la fois aux élèves et à l'établissement scolaire.

L'expérience menée par McIntyre *et al.* (2005) auprès d'une quarantaine d'élèves entrant dans le secondaire démontrent une cohérence des résultats autour de plusieurs points : de la part des élèves, l'expression d'**un relatif consensus sur ce qui les aiderait dans leur apprentissage, focalisé en particulier sur la dimension sociale de l'apprentissage (plus d'interactivité, plus de travail collaboratif et plus d'authenticité), et des suggestions fortement inspirées de ce qu'ils connaissent déjà** ; de la part des enseignants, une plus grande divergence dans la prise en compte de ces suggestions, bien que considérées comme pertinentes, dévoilant des niveaux de confort ou d'inconfort variables, et un engagement à plus ou moins long terme dans la rénovation de leurs pratiques.

L'enquête réalisée en Irlande du Nord auprès d'élèves de classes dites « formatives » (*AfL classrooms*) dans le cadre du projet CPAL (Consulting Pupils on the Assessment of their Learning) apporte un éclairage complémen-

taire convergent (Leitch & Odena, 2007). Selon les auteurs, cette « voix de l'élève » peut être chronologiquement décomposée en quatre facteurs : il doit avoir l'opportunité d'exprimer un point de vue (*space*), cette prise de parole doit être facilitée, encouragée (*voice*) et bénéficier de l'attention d'un public (*audience*), avant d'être prise en considération de façon appropriée (*influence*). D'après les conclusions de l'enquête, les élèves reconnaissent que les opportunités de s'exprimer existent (*space* et *voice*), mais restent plus circonspects sur les deux autres facteurs, estimant que leur feedback n'est pas suffisamment écouté et quasiment jamais suivi d'actions. Les enseignants pour leur part, trouvent les retours des élèves utiles, en particulier pour mieux comprendre leurs propres pratiques. Mais ils considèrent l'organisation spatio-temporelle de l'école, les contraintes curriculaires et l'obligation de résultats à laquelle ils sont soumis, comme autant de freins à la prise en compte effective de cette « voix de l'élève ». De surcroît, le besoin de contrôler la classe, le manque de recul en situation, la culture scolaire et le type de matières enseignées jouent également un rôle dissuasif important pour franchir l'étape de la régulation des enseignements.

Intégrer la consultation des élèves en quelque sorte à la routine pédagogique dans une perspective régulative soulève un certain nombre de questions non résolues par les expériences relatées. Hoban et Hastings (2006) ont démontré que parmi les quatre méthodes de collecte examinées, l'entretien personnel est la plus susceptible d'influencer les pratiques pédagogiques de l'enseignant. Un tel mode de consultation des élèves non seulement ne permet pas d'écouter toutes les voix, mais paraît difficilement généralisable. **Faut-il considérer que l'hypothèse régulative est mal posée et que cette voix de l'élève, notamment au vu des freins liés à la potentielle concurrence entre programmes et pédagogie, ne peut bénéficier que d'une prise en compte non formelle ?** Force est de conclure avec Michael Fielding, dans un numéro spécial de la revue *Discourse: Studies in the Cultural Politics of Education* (2007) consacré à la série de séminaires intitulée Engaging Critically with Pupil Voice (2004-2006), que le caractère émergent de ce champ scientifique ne permet guère d'apporter de réponses claires pour l'instant.

L'évaluation des compétences

Quelle évaluation pour quelle compétence ?

Dans le vocabulaire courant de l'évaluation, le terme de compétences peut aussi bien servir à qualifier une connaissance ou un savoir-faire qu'à désigner une approche complexe et intégrée qui s'oppose à la décomposition en objectifs détaillés et fragmentés.

À partir des évaluations nationales « diagnostiques » de CE2 et d'entrée en sixième, Sophie Morlaix a identifié les « compétences » pour la réussite scolaire (calcul mental, par exemple), en dégageant des regroupements d'items de réponse les plus prédictifs de la réussite de l'élève (Morlaix, 2007). Si ces regroupements permettent effectivement d'identifier des compétences élémentaires d'apprentissage au-delà des savoirs disciplinaires constitués, ils ont peu à voir avec les « compétences » complexes visées par d'autres chercheurs ou réformes. On peut dire, en la matière, que le *Socle commun de connaissances et de compétences* français mélange les deux conceptions, et que l'évaluation à travers les « livrets de compétences » reste problématique faute d'une clarification de ce que l'on entend par « compétences » et d'une explicitation des fonctions que l'on souhaite faire jouer aux divers outils d'évaluation (Houchot & Robine, 2007).

Sans revenir plus longuement sur cette question générale de l'approche par compétences, abordée notamment dans un précédent dossier de l'INRP (Rey, 2008), on s'intéressera aux relations entre compétences et évaluation, considérée dans sa version la plus « complexe ». Le moment de l'évaluation constitue en effet le point le plus problématique de l'approche par compétences, puisqu'il s'agit dans ce cas d'évaluer non plus seulement des savoirs scolaires en situation scolaire, mais de mobiliser dans une situation inédite, et parfois proche de la vie « réelle », des ressources, habiletés, compétences, etc.

Pour Scallon, l'un des spécialistes de l'évaluation des compétences (voir en particulier son site pédagogique), il faut ainsi distinguer la situation d'évaluation de niveau **connaissance**, qui est plutôt associée à la mémorisation d'informations, et la situation de niveau **habiletés**, dans laquelle l'individu doit utiliser ses connaissances pour élaborer une réponse qu'il n'a pas apprise par cœur au préalable (Scallon, 2007, p.33-46), la situation de connaissances intégrant les deux précédentes.

Une situation de connaissance se caractérise par la demande directe d'une information que l'élève doit avoir mémorisée et doit posséder dans son répertoire cognitif. Dans une situation d'habileté, l'élève doit se demander quelles connaissances utiliser mais l'objet de la question et le domaine dans lequel elle s'inscrit sont connus (situation scolaire balisée), alors que dans la situation de compétence, il faut identifier le type de domaines et les connaissances et habilités pertinentes à mobiliser pour répondre à une question posée dans des termes pas forcément scolaires.

La notion de distance (temporelle et conceptuelle) entre la situation d'apprentissage et la tâche demandée est donc importante pour apprécier l'évaluation des compétences. La question du développement d'un processus d'évaluation intégré mais qui se déploie dans toute la durée du temps de l'apprentissage (et couvre l'ensemble du champ) est aussi jugée primordiale par Tardif, qui insiste sur le fait que l'évaluation des compétences est à l'inverse de la juxtaposition d'une série de « tests » conçus isolément les uns des autres. L'évaluation des compétences doit donc être étalée dans le temps pour que l'on puisse en cerner le développement (Tardif, 2006).

Quant à la différenciation entre compétences proches des savoirs élémentaires et compétences plus larges, des chercheurs belges qui ont travaillé de façon approfondie sur la question de l'évaluation des compétences (Rey, Carette, Defrance & Kahn, 2006) distinguent trois niveaux dont seuls les deux derniers méritent vraiment, à leurs yeux, d'être appelés compétences :

- la compétence élémentaire : savoir exécuter une opération en réponse à un signal (procédure automatisée, habileté) ;
- la compétence avec cadrage : interpréter une situation inédite et choisir la compétence élémentaire qui convient ;
- la compétence complexe : choisir et combiner plusieurs compétences pour traiter une situation nouvelle et complexe.

L'objectif des auteurs est d'abord d'évaluer des compétences complexes, autrement dit, la capacité des élèves à choisir et à combiner, parmi les procédures qu'ils ont déjà « apprises », plusieurs d'entre elles, afin de résoudre de façon adéquate un problème nouveau pour eux. Ce qui est recherché dans ces « modèles d'évaluation », c'est **l'autonomie intellectuelle plus que l'automatisme**.

Pour que l'évaluation garde son caractère de construction de l'apprentissage, elle ne doit pas être uniquement sommative mais doit avoir un caractère diagnostique. L'épreuve doit aussi comprendre une seconde phase, où la tâche complexe est décomposée en procédures de base à choisir (tâches partielles). Enfin une troisième phase de vérification de la maîtrise décontextualisée des procédures de base (tâches simples), peut être nécessaire pour vérifier, à l'issue du dispositif, à quel stade se situent les difficultés rencontrées.

Ainsi, il est possible de déterminer si l'élève bute sur la résolution de la situation nouvelle et complexe (identification du problème, choix et mobilisation des procédures pertinentes), sur le choix de la procédure adaptée ou sur la maîtrise même des procédures de base ou des compétences élémentaires.

À partir de cette démarche et des enquêtes pratiques qui s'en sont inspirées, V. Carette a par ailleurs avancé l'hypothèse que les épreuves d'évaluation centrées sur les compétences n'évaluaient pas le même type d'acquis que des évaluations plus classiques, centrées sur les résultats de l'enseignement par objectifs, ce qui interpelle quant à la pertinence de juger de l'efficacité des différentes approches pédagogiques sans interroger le type d'évaluation utilisé (Carette, 2008).

Évaluer une compétence transversale : exemple du savoir-apprendre

La question de savoir comment développer les compétences individuelles relatives à l'apprentissage tout au long de la vie, et donc comment mesurer les progrès en lien avec ces compétences, est inscrite dans l'agenda européen depuis les premières réflexions qui ont nourri la stratégie de Lisbonne en 2000. Mais passer d'un système qui évalue des connaissances à un système qui évalue (aussi) des compétences implique un changement de perspective qui place les acteurs des systèmes éducatifs en difficulté. Cette tendance à vouloir articuler ce qui est traditionnellement du ressort de l'institution scolaire (acquisition de connaissances et développement de capacités cognitives – *hard* ou *cognitive outcomes*) à des enjeux explicites liés au développement personnel, social et moral de la personne (*soft* ou *affective outcomes*), génère des questionnements inédits sur les responsabilités respectives des différents acteurs et les moyens à mettre en œuvre.

La création du Centre for Research on Education and Lifelong Learning (CRELL) en 2005 traduit la volonté politique européenne de rendre opérationnelle au sein des systèmes éducatifs cette approche par compétences. Les premiers travaux donnent lieu en décembre 2006 à l'adoption par le Parlement européen et le Conseil de l'Europe d'une recommandation qui identifie la compétence « apprendre à apprendre » comme l'une des huit compétences clés pour l'éducation et la formation tout au long de la vie. Remarquons que l'OCDE la considère comme une compétence transverse, constitutive des trois catégories de compétences clés de son projet DeSeCo (2002) : « *interagir dans des groupes hétérogènes, agir de façon autonome, se servir d'outils de manière interactive* ». De même, le Socle commun de connaissances et de compétences en France (2006) ne compte pas le savoir-apprendre au rang des sept compétences clés qu'il propose, malgré un écho évident au niveau de la septième compétence « autonomie et esprit d'initiative ».

Ce qui nous intéresse présentement ne concerne pas tant les débats sur la position du savoir-apprendre dans le curriculum, que les propositions émises par un groupe d'experts du CRELL visant à élaborer un test européen pour mesurer cette même compétence (Hoskins & Fredriksson, 2008). Selon eux, la base fournie par les questions relatives à la résolution de problèmes dans PISA 2003 (abandonnées dans PISA 2006 et 2009 mais possiblement réintroduites dans l'édition 2012) n'offre qu'une approche indirecte du savoir-apprendre. L'ambition du test du CRELL porte sur une approche interdisciplinaire et authentique, intégrant les dimensions à la fois affective, cognitive et méta-cognitive de cette compétence.

Pour construire cette définition formelle et concevoir l'outil de mesure adapté, les experts se sont appuyés sur les expériences existantes : le test Effective lifelong learning inventory (ELLI) de l'université de Bristol pour les aspects socioculturels, le projet finlandais The L2 factor. Learning-to-Learn at School – A Key to Lifelong Learning à l'université d'Helsinki et le test de compétences transversales (*cross-curricular skills test* ou CCST) de l'université d'Amsterdam pour les aspects cognitifs, et enfin les travaux de la faculté de psychologie de l'université autonome de Madrid sur la mesure des capacités méta-cognitives.

À l'issue de la phase pilote (2300 élèves de 14 ans dans une cinquantaine d'écoles en Europe), les rapports nationaux soulignent la difficulté à considérer séparément les dimensions affective et cognitive de l'apprentissage, et relèvent les différences culturelles dans les réponses des élèves. Ce travail préliminaire s'oriente ainsi vers un renforcement de la démarche interdisciplinaire et donc vers une **articulation plus forte entre une définition formelle et normative de la compétence d'une part et une multiplicité de contextes sollicitant cette compétence d'autre part**.

Une autre piste explorée dans un rapport complémentaire du CRELL (Hoskins & Deakin Crick, 2008), est celle d'un rapprochement entre les indicateurs du savoir-apprendre et ceux liés aux compétences civiques, également comptées parmi les huit compétences clés pour l'éducation et la formation tout au long de la vie (2006). Un tel rapprochement fait également écho aux travaux sur l'impact social de l'éducation (par exemple CERI, 2007) qui défendent la thèse d'une corrélation entre le niveau d'éducation et le bien-être social, et qui se traduisent notamment par l'émergence de programmes d'» éducations à… » (la santé, la citoyenneté, etc.) dans les curriculums.

Dans le rapport du CRELL, Hoskins et Deakin Crick croisent analyses conceptuelles et données empiriques pour mettre en évidence les similitudes entre ces deux compétences, et en particulier leur importance pour agir dans la société : **le savoir-apprendre pour devenir un apprenant actif et les compétences civiques pour devenir un citoyen actif**. Considérant l'une et l'autre comme complémentaires, ils suggèrent d'explorer davantage les initiatives et dispositifs existants, comme par exemple les travaux de CitizED, communauté d'intérêt anglaise, dédiée à la formation initiale et continue des enseignants sur les questions d'éducation à la citoyenneté.

Car malgré des points de convergence avec le projet anglais LHTL cité précédemment, les travaux du CRELL privilégient une évaluation macro à l'échelle de l'Europe, dont les finalités – diagnostiques, sommatives ou politiques ? – demeurent incertaines. Le fait que les niveaux micro (classe) et méso (établissement) au cœur de l'approche formative ne soient pas ici sollicités, relance **la question du « qui évalue ? » et en particulier du « qui évalue les compétences ? »**. Si dans le cadre du projet LHTL le savoir-apprendre est perçu comme constitutif de l'apprentissage et relève donc de la responsabilité de l'enseignant, le CRELL opte pour une mesure instrumentée ponctuelle, interdisciplinaire et extérieure à l'acte d'apprentissage : une vision qui s'inscrit notamment dans la continuité des initiatives du Centre for educational assessment (CEA) de l'université d'Helsinki et qui renouvelle le genre des tests (inter)nationaux, généralement peu enclins à prendre en compte l'individu derrière l'élève ni l'environnement d'apprentissage. La seule exception notable est sans doute l'enquête internationale de l'IEA sur l'éducation civique et la citoyenneté – International Civic and Citizenship Education Study (ICCS), inspirée d'une précédente expérience conduite en 1999 (Study of civic education ou CIVED) – dont la première édition sera administrée en 2009 dans une quarantaine de pays.

Ce choix d'investir dans un programme d'évaluation externe semble répondre davantage à des exigences politiques relatives à une certaine obligation de résultats (*accountability*) qu'à une logique d'instrumentation au service de la régulation des apprentissages, quand bien même les jalons posés par la définition de l'instrument représenteraient une étape dans le développement de l'approche par compétences. Ce choix n'exclut donc pas a priori une inscription explicite de ces compétences dans la pédagogie, telle qu'elle peut être soutenue par exemple par les travaux du Citizenship Education Research Strategy Group de l'EPPI-Centre londonien : les deux revues de littérature dirigées en 2004 et 2005 par R. Deakin Crick ont ainsi montré l'efficacité de pédagogies centrées sur l'élève, dans un environnement scolaire qui privilégie la qualité du dialogue et la confiance d'une part, et un ancrage pédagogique dans la vie réelle d'autre part, pour développer le savoir-apprendre et les compétences civiques. Ces synthèses ont également mis en évidence les bénéfices de cette intégration en termes de résultats scolaires, confortant parallèlement des initiatives telles qu'ELLI qui ambitionne de construire un instrument pour mesurer les compétences individuelles en matière d'apprentissage tout au long de la vie.

Depuis 2002, les travaux d'ELLI, coordonnés par Ruth Deakin Crick, Patricia Broadfoot et Guy Claxton (université de Bristol), se sont concentrés sur le concept clé de « pouvoir apprendre » (*learning power*), ainsi défini : « *un mélange complexe de dispositions, d'expériences, de relations sociales, de valeurs, d'attitudes et de convictions qui fusionnent pour former la nature de l'engagement de l'individu au moment où une opportunité d'apprentissage se présente* ».

Les sept dimensions constitutives de ce pouvoir apprendre (*changing & learning, meaning making, curiosity, creativity, learning relationships, strategic awareness and resilience*), ont elles-mêmes permis de différencier grossièrement deux types d'apprenants : des élèves efficaces, engagés et dynamiques d'une part, des élèves passifs, dépendants et fragiles d'autre part (*efficacious, engaged and energised learners and passive, dependent and fragile learners*) (Deakin Crick, 2007).

L'utilisation de l'outil de mesure élaboré dans le cadre d'ELLI (aujourd'hui mis à la disposition des établissements scolaires) a clairement montré que plus les élèves avancent dans leur scolarité, plus ils deviennent faibles dans l'ensemble des sept dimensions identifiées, et en particulier pour ce qui concerne la créativité ; autrement dit, **plus ils deviennent fragiles et dépendants dans leurs apprentissages**. À l'aide des profils d'apprentissage détaillés de leurs élèves, les enseignants ont développé de nouveaux modes d'intervention dans la classe, depuis une réorganisation de la manière d'enseigner à un focus sur des pratiques d'autoévaluation. Les bénéfices se sont avérés mesurables rapidement : après seulement deux trimestres, les élèves

se sont montrés plus capables de résister au jeu scolaire de l'évaluation sommative et plus conscients de leurs propres stratégies d'apprentissage. Selon les auteurs, les facteurs clés de cette évolution résident à la fois dans les représentations de l'enseignant et dans le **climat scolaire**.

Le portfolio : un outil réflexif au service de l'évaluation ?

Une démarche plutôt qu'un outil

La question de savoir ce qu'est un portfolio et à quoi il sert reste confuse pour une majorité des acteurs de l'éducation, tant les représentations sont brouillées par la multiplicité des contextes dans lesquels ils sont utilisés (de l'école primaire à la formation des adultes), par les multiples fonctions qu'on lui assigne (évaluation formative, validation des acquis, etc.) et par les divers univers conceptuels qu'ils sollicitent (psychologie cognitive, usage des TIC, etc.). La France n'échappe pas à cette confusion, cumulant dans les textes officiels la mention plus ou moins prescriptive de supports connexes aux appellations variées : livret scolaire pour l'école élémentaire, livret de l'apprenti ou livret d'apprentissage pour l'enseignement professionnel, portfolio pour les compétences en langues vivantes, livrets de compétences pour l'évaluation des acquis, dossier scolaire pour l'orientation, etc. (Houchot & Robine, 2007).

Dans le domaine de l'éducation et de la formation, ce sont historiquement les études médicales et la formation des enseignants qui bénéficient d'une plus longue tradition en matière d'usage de portfolios (Butler, 2006), alors que les expériences dans l'enseignement scolaire sont plus récentes (voir aussi le dossier d'Educnet et les Cent références pour le portfolio numérique de Robert Bibeau, l'un et l'autre mis à jour en 2008).

Une des définitions couramment citée est celle de Paulson *et al.* qui considèrent le portfolio comme « *une collection significative des travaux de l'élève illustrant ses efforts, ses progrès et ses réalisations, dans un ou plusieurs domaines* » (Paulson *et al.* 1991, cité et traduit par Goupil, 1998).

Cette définition a progressivement évolué, notamment sous l'influence des supports numériques. Celle fournie par Challis (2005), auteur de travaux sur l'usage des portfolios dans la formation médicale, est aujourd'hui 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, et intégrées à un site web ou disponible en cd-rom ou dvd-rom* ».

Les questionnements autour des outils sont marqués par des préoccupations de pérennité (standards, interopérabilité, etc.) et de continuité pédagogique, liée en particulier aux moments clés de transition dans le parcours de l'élève (Cloutier *et al.*, 2006). Selon George Siemens toutefois (2004), **la standardisation est préjudiciable à la prise en main du système par les élèves et répond davantage à des logiques institutionnelles, voire politiques, qu'à une démarche pédagogique.** Autrement dit, l'utilisation d'un outil simple au départ (logiciel de présentation, traitement de texte, blog à un premier niveau, ou plate-forme dynamique modulaire de type CMS à un 2e niveau) est essentielle pour accélérer l'adoption de la démarche et permettre à l'élève de prendre ses décisions et de garder le contrôle sur le contenu de son portfolio et l'accès à ce contenu.

Dans sa revue de littérature, Butler insiste également sur la nécessaire flexibilité et modularité de l'outil (Butler, 2006), et sur l'équilibre à trouver entre une interface fonctionnellement limitée et un outil trop compliqué et peu intuitif. Le rapport publié dernièrement dans le cadre du projet européen More self esteem with my e-Portfolio (MOSEP), s'inspire de cette hypothèse pour offrir un comparatif détaillé des solutions de portfolio open source existantes, permettant ainsi d'évaluer leur adéquation à des utilisateurs débutants (Attwell *et al.*, 2007). L'étude du BECTA (Hartnell-Young, 2007) met également en évidence cette **tension entre la créativité des élèves et la sophistication de l'architecture de l'outil utilisé**, et souligne le potentiel des outils sociaux externes à l'environnement scolaire formel. À la lumière des études de cas analysées, les auteurs insistent sur l'importance de la dimension sociale : l'outil doit permettre le feedback de tiers, enseignants et pairs (commentaires, espaces de discussion, etc.) pour que la démarche gagne en efficacité. Ce discours sur la flexibilité nécessaire de l'outil se traduit aussi par l'émergence d'un concept connexe, celui d'**espace numérique personnel** (*personal learning environment*), garantissant à chacun la possibilité d'archiver les traces de ses activités et participant à la construction de son **identité numérique**, comme l'explique Stephen Downes, conférencier invité au colloque Eportfolio 2008 de Montréal (Downes, 2008).

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. Helen Barrett (2004) par exemple découpe le processus en cinq phases : collecter (*collecting*), sélectionner (*selecting*), réfléchir (*reflecting*), montrer ou partager (*projecting*) et publier ou rendre publique 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*) (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.

Finalité unique ou multiple ?

Un portfolio, oui, mais pour faire quoi ?… La question des finalités mixtes est au cœur de tous les questionnements : développement personnel ? apprentissage ? évaluation ? présentation ? Ces finalités, apparemment distinctes, peuvent-elles cohabiter sur un même support ?

En s'appuyant sur la littérature existante, Helen Barrett (2004) conclut que l'utilisation d'un portfolio pour des évaluations à forts enjeux (*high stakes assessment of learning*) diffère fondamentalement des usages à visée formative dans le cadre d'une évaluation pour l'apprentissage. Dans le premier cas, la collecte des preuves s'organise par rapport à des attentes externes standardisées et l'objectif est de mesurer ce qui a été appris dans un cadre temporel prédéfini : l'évaluateur s'intéresse donc au produit. Dans le second, la collecte, éventuellement négociée avec l'enseignant, relève en priorité du choix de l'élève, et l'objectif est de lui permettre de « raconter » son expérience d'élève pour évaluer ses besoins à venir : ici le focus porte sur le processus. Si un portfolio peut servir l'une et l'autre finalité, l'auteur défend l'idée que le concept même de portfolio est potentiellement plus adapté à une perspective formative, et elle établit un parallèle entre l'usage d'un portfolio et l'acte d'écrire sur soi (*storytelling*).

Parmi ces nombreuses acceptions et utilisations, Gérard Scallon (2007) considère également que le principal intérêt du portfolio réside dans l'auto-évaluation. Dans cette perspective, il distingue trois types d'usages, 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.

Si l'on retrouve dans son analyse cette dimension narrative mentionnée par H. Barrett, Scallon va plus loin, en postulant 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).

En revanche, ces dossiers doivent **se différencier clairement des bulletins ou livrets scolaires**. L'expérience suisse est éloquente sur ce point (Gilliéron Giroud, 2007). Le dossier d'apprentissage, introduit en Suisse romande dans les années 90, a progressivement endossé une fonction de communication, au point de devenir dans le Canton de Vaud le seul support transmis aux familles justifiant des décisions institutionnelles (passage au niveau supérieur, orientation, etc.). Cette multiplication des finalités et des destinataires a constitué un frein important dans la phase de généralisation du dispositif, générant de nombreuses incompréhensions de part et d'autres, et s'est traduit par un retour aux notes et moyennes, précédemment abandonnées.

Val Klenowski, auteur d'un des rares ouvrages intégralement dédiés aux portfolios (2002), exprime des réticences similaires à celles d'H. Barrett, s'agissant de l'utilisation d'un portfolio pour la certification des enseignants. Selon elle, ce qui importe dans la définition du portfolio n'est pas tant la collection de travaux qu'il contient que sa « position centrale dans l'auto-évaluation et la réflexion des élèves, ainsi que l'opportunité de décrire le processus par lequel le travail dans le portfolio est réalisé » (p. 3). Illustrant son propos avec des exemples pris dans des contextes très divers (depuis l'enseignement primaire et secondaire jusqu'à la formation initiale et continue des enseignants et des médecins, depuis l'Australie jusqu'à l'Ecosse, en passant par Hong Kong, etc.), l'auteur met en évidence les corrélations entre portfolio et évaluation pour l'apprentissage. Elle souligne l'existence d'une culture commune centrée sur le développement métacognitif de l'élève, composée d'éléments tels que la compréhension des critères d'évaluation par les élèves, leur engagement dans le processus d'évaluation et l'importance d'une approche intégrée de l'évaluation, allant jusqu'à l'évaluation du portfolio en tant que produit (*grading of the portfolio*). Selon elle, le portfolio peut ainsi répondre à deux types de finalités : l'élève est tenu de faire la preuve de ses apprentissages, soit dans une perspective formative (focus sur le processus d'apprentissage), soit dans une perspective sommative (focus sur la maîtrise de certaines compétences), soit dans une combinaison des deux. Autrement dit, **le portfolio réflexif peut rendre compte à la fois des résultats de l'apprentissage et de la qualité du processus d'apprentissage lui-même**.

Des pratiques inégales

Les expériences dont fait état l'abondante littérature sur les portfolios permettent de dégager quelques invariants en termes de facteurs prédictifs de succès. Tout d'abord le fait que l'élève se sente **propriétaire de son portfolio**, c'est-à-dire responsable des documents collectés, de leur organisation, voire de leur présentation, est considéré comme crucial par une majorité d'observateurs. Corrélativement, le degré de flexibilité du portfolio est un prérequis à une intégration pédagogique de la démarche, comme évoqué précédemment. Cette flexibilité se traduit notamment par la diversité des éléments qui peuvent être collectés pour servir de preuves, et **par la diversité des stratégies d'apprentissage** que le travail avec le portfolio permet d'émuler ; il s'agit

par exemple de ne pas se limiter à ce que l'élève dit qu'il a fait, en incluant le cas échéant des traces vidéos, en offrant plusieurs méthodes pour organiser et trier les preuves collectées, ou bien encore en promouvant une validation à la demande, quand l'élève est prêt. (Attwell *et al.*, 2007 ; Butler, 2006 ; Hartnell-Young, 2007 ; Schärer, 2007).

L'expérience belge conduite auprès d'élèves du premier degré du secondaire (Bernard & Vlassis, 2007) met en évidence une plus forte motivation des élèves et des enseignants participants quand plusieurs disciplines sont impliquées dans le projet. Le fait que les usages des portfolios soient inégalement répandus selon les environnements professionnels considérés conduit à s'interroger sur les affinités potentielles entre certaines disciplines, plus expérimentales ou plus professionnelles, et la démarche induite par le portfolio. Les investigations menées dans 35 environnements disciplinaires différents, dans le cadre du projet ISLE (Individualised support for learning through eportfolios), infirment cette hypothèse, en montrant que **l'efficacité n'est pas liée au contexte disciplinaire** (ISLE, 2007).

Si la démarche paraît plus facile à implémenter à l'école primaire du fait d'une organisation plus flexible (Houchot & Robine, 2007), Louise Bélair montre toutefois que l'organisation du portfolio par discipline reste dominante ; il remplit tout au plus une fonction d'auto-régulation, mais ne rend pas compte d'un niveau de compétences et n'est donc pas formellement utilisé dans un contexte d'évaluation. Il rencontre toutefois l'adhésion conjointe des élèves et des enseignants interrogés, considérant les activités générées par le portfolio comme une forte source de motivation. Sa vocation plus « sommative » n'interviendrait-elle donc que dans un second temps ? (Bélair, 2007).

Rien n'est moins sûr si l'on se réfère à l'expérience suisse, qui s'est soldée par l'exclusion de toute fonction d'évaluation formelle pour privilégier la fonction de métacognition du portfolio. Patricia Gilliéron Giroud rend compte de nombreuses difficultés rencontrées par les enseignants et pointe les faiblesses relatives à l'introduction des portfolios dans les classes de Suisse romande (Gilliéron Giroud, 2007) : à la confusion générée par ses multiples finalités, s'ajoutent la lourdeur du processus liée à la personnalisation de l'approche, les interrogations sur la validité des preuves collectées par l'élève et **la difficulté à analyser les traces d'apprentissage pour qualifier des progrès**.

En tout état de cause, la nature et l'importance des interventions des enseignants et des pairs dans le processus de construction du portfolio, son inscription à plus ou moins long terme dans les pratiques (quand un portfolio est-il fini ?) et la manière dont il est rendu public (quels sont les destinataires ?) sont au cœur des interrogations pédagogiques (Butler, 2006). Comme pour tous les outils d'évaluation ouverts, un autre enjeu important est d'apporter des « **clés de fermeture** » susceptibles de réduire la part de subjectivité voire d'arbitraire. Dans l'évaluation des compétences par situations complexes via des portfolios, la connaissance par les apprenants des critères sur lesquels ils seront évalués est donc essentielle (Gérard, in Baillat *et al.*, 2008).

Si de nombreuses études tendent à montrer l'impact du portfolio sur la motivation des élèves et postulent son influence sur les résultats scolaires, les questions de mise en œuvre restent également insuffisamment éclairées. Car malgré les freins, les prescriptions institutionnelles existent : généralisation du PEL (portfolio européen des langues) dans les écoles primaires et les collèges suisses d'ici 2010, généralisation du portfolio d'évaluation pour le diplôme de fin d'études secondaires en Colombie Britannique, etc. Dès lors, comment créer des conditions favorables au déploiement d'un projet de portfolio ? Les injonctions à réformer la culture scolaire, à réorganiser programmes et emplois du temps, à intégrer le portfolio dans le quotidien des apprentissages, etc. ne suffisent pas.

Certains contextes sont évidemment plus favorables que d'autres : l'étude du BECTA (Hartnell-Young, 2007) met par exemple en évidence les corrélations fortes entre e-portfolios et TIC, montrant que le portfolio numérique est d'autant plus facile à promouvoir dans un établissement où les TICé préexistent. Tous les observateurs insistent cependant sur la planification d'une phase pilote suffisamment longue, permettant d'étudier la faisabilité concrète de la généralisation dans les classes, de définir les structures, centralisées ou décentralisées, chargées d'accompagner et de soutenir le développement du dispositif (conseil, formation continue, etc.) avec les équipes pédagogiques, et pour finir de déterminer un calendrier raisonnable et des indicateurs réalistes (voir en particulier Schärer, 2007). Dans cette démarche d'introduction, les travaux menés au niveau européen peuvent accompagner les réflexions.

Des initiatives européennes : PEL et MOSEP

Le Portfolio européen des langues (PEL) (*European language portfolio* ou ELP), élaboré par le Conseil de l'Europe, incarne sans doute l'une des expériences les plus extensives de portfolio à l'échelle européenne. Lancé en 2001 à l'occasion de l'année européenne des langues, le PEL est développé dans vingt-huit pays membres et comprend fin 2008 près d'une centaine de modèles validés, couvrant tous les secteurs de l'enseignement, depuis l'enseignement primaire jusqu'à la formation des adultes (voir aussi les études de cas rassemblées par David Little en 2003).

Le rapport intérimaire 2006 (Schärer, 2007) fait état de résultats positifs significatifs à la fois comme outil pédagogique au service du développement de l'autonomie et comme outil de présentation témoignant des compétences langagières. Mais les expériences en cours n'ont pas encore atteint un niveau de maturité suffisant pour permettre d'analyser réellement l'impact du portfolio sur les résultats d'apprentissage. Le fait que le PEL **valo-**

rise cependant les « réussites » dans une perspective à plus ou moins long terme et dans un cadre dépassant l'horizon scolaire contribue fortement à l'adhésion des élèves.

Une autre caractéristique clé du PEL est sa prise en compte explicite de l'ensemble des compétences langagières de l'élève, « qu'elles soient apprises ou acquises dans le cadre du système éducatif officiel ou en dehors de celui-ci ». Cette **reconnaissance de ce que l'élève sait, indépendamment de l'espace-temps scolaire**, est également une composante fondamentale d'un autre projet européen, le MOSEP (2006-2008), placé sous la direction du Salzburg Research Forschungsgesellschaft (Attwell *et al.*, 2007). Ce projet, intitulé More self esteem with my e-Portfolio propose une boîte à outils et un ensemble de tutoriels guidant enseignants et conseillers d'orientation à mettre en œuvre un portfolio avec des adolescents en situation de décrochage à la fin de la scolarité obligatoire. La méthodologie déployée s'enracine dans les travaux de Bandura et Pajares sur la psychologie adolescente. L'objectif est donc clairement de mettre en œuvre un accompagnement personnalisé de l'élève, via l'usage d'un portfolio, pour accroître sa motivation, en misant sur son sentiment d'efficacité personnelle (*self efficacy*) et sa capacité à réguler ses propres apprentissages (*self-regulated learning*). Le projet écossais ISLE – Individualised Support for Learning through e-Portfolios – soutient l'hypothèse fondatrice de MOSEP, démontrant que le portfolio peut atteindre plus d'efficacité quand il est **associé à des moments clés du parcours des apprenants** (ISLE, 2007).

Bibliographie

- (2007). « Beyond "Voice": New roles, relations, and contexts in researching with young people ». *Discourse: Studies in the Cultural Politics of Education*, vol. 28, n° 3, p. 301-420.
- (2007). *ISLE: Individualised support for learning through e-Portfolios*. Joint Information Systems Committee (JISC).
- (2006). « Learning How to Learn, in Classrooms, Schools and Networks ». *Research Papers in Education*, vol. 21, n° 2, p. 101-234.
- Agrawal Manta (2004). « Curricular reform in schools: the importance of evaluation ». *Journal of Curriculum Studies*, vol. 36, n° 3, p. 361-379.
- Allal Linda (2007). « Régulation des apprentissages : Orientation conceptuelle pour la recherche et la pratique en éducation ». In Allal Linda & Mottier Lopez Lucie (dir.). *Régulation des apprentissages en situation scolaire et en formation*. Bruxelles : De Boeck, p. 7-23.
- 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 & Lafortune Louise (dir.) (2008). *Jugement professionnel en évaluation : Pratiques enseignants au Québec et à Genève.* Sainte-Foy : Presses de l'université du Québec.
- Attwell Graham, Chrzaszcz Agnieszka & Pallister John *et al.* (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*. Salzburg : Salzburg Research Forschungsgesellschaft.
- Baker Eva L. (2007). « 2007 Presidential Address-The End(s) of Testing ». *Educational Researcher*, vol. 36, n° 6, p. 309-317.
- Barrett Helen (2004). « Electronic Portfolios as Digital Stories of Deep Learning ». *Using Technology to Support Alternative Assessment and Electronic Portfolios*. En ligne (consulté le 13 novembre 2008) : <http://electronicportfolios.org/digistory/epstory.html>.
- 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.
- Bélair Louise (2007). « Le dossier de l'élève (portfolio) dans l'approche par compétences au primaire au Québec : Quels apports ? quelles perspectives ? quelles limites ? ». In *Actes du 19e colloque de l'ADMEE-Europe*, Université du Luxembourg, 11-13 septembre 2006.
- Berliner David C. & Amrein Audrey L. (2002). « High-Stakes Testing, Uncertainty, and Student Learning ». *Education Policy Analysis Archives*, vol. 10, n° 18.
- Bernard Hervé & Vlassis Joëlle (2007). « Le portfolio au service de l'enseignement des compétences transversales ». In *Actes du 19e colloque de l'ADMEE-Europe*, Université du Luxembourg, 11-13 septembre 2006.
- Black Paul & Wiliam Dylan (1998). « Assessment and Classroom Learning ». *Assessment in Education*, vol. 5, n° 1, p. 7-68.
- Brennan John & Williams Ruth (2004). *Collecting and using student feedback: A guide to good practice*. Bristol : Higher Education Funding Council for England (HEFCE).
- Bressoux Pascal & Pansu Pascal (2003). *Quand les enseignants jugent leurs élèves*. Paris : Presses universitaires de France.
- Broadfoot Patricia & Pollard Andrew (2006). « The Changing Discourse of Assessment Policy: The Case of English Primary Education ». In Lauder Hugh, Brown Phillip, Dillabough Jo-Anne & Halsey A. H. (dir.). *Education, Globalization & Social Change*. New York : Oxford University Press.
- Broadfoot Patricia (2007). *An Introduction to Assessment*. London : Continuum.
- Butler Philippa (2006). *A Review Of The Literature On Portfolios And Electronic Portfolios*. Palmerston North, New Zealand : Massey University College of Education.

- Carette Vincent (2008). « Les caractéristiques des enseignants efficaces en question ». *Revue française de pédagogie*, n° 162, p. 81-93.
- Cariola M. Leonor & Meckes Lorena (2008). « L'évolution du rôle du système d'évaluation : Le cas du Chili ». *Revue internationale d'éducation – Sèvres*, n° 48, p. 37-51.
- Cédelle Luc (2008). *Un plaisir de collège*. Paris : Le Seuil.
- 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.
- Centre pour la recherche et l'innovation dans l'enseignement (CERI) (2007). *Comprendre l'impact social de l'éducation*. Paris : OCDE.
- Challis Diana (2005). « Towards the mature ePortfolio: Some implications for higher education ». *Canadian Journal of Learning and Technology*, vol. 31, n° 3.
- Cloutier Martine, Fortier Guy & Slade Samantha (2006). *Le portfolio numérique : Un atout pour le citoyen apprenant*. Québec : Société de formation à distance des commissions scolaires du Québec (SOFAD).
- Crooks Terence J. (1988). « The Impact of Classroom Evaluation Practices on Students ». *Review of Educational Research*, vol. 58, n° 4, p. 438-481.
- Cytermann Jean-Richard & Demeuse Marc (2005). *La lecture des indicateurs internationaux en France : rapport et avis*. Paris : La Documentation française.
- Dauvisis Marie-Claire (2007). « L'évaluation des compétences au risque des barèmes et des notes scolaires ». In Hatano Maude & Lemaître Denis (dir.). *Usages de la notion de compétence en éducation et formation*. Paris : L'Harmattan, p. 75-92.
- Deakin Crick Ruth (dir.) (2004). *A systematic review of the impact of citizenship education on the provision of schooling*. London : EPPI-Centre.
- Deakin Crick Ruth (dir.) (2005). *A systematic review of the impact of citizenship education on student learning and achievement*. London : EPPI-Centre.
- Deakin Crick Ruth (2007). « Learning how to learn: The dynamic assessment of learning power ». *The Curriculum Journal*, vol. 18, n° 2, p. 135-153.
- Downes Stephen (2008). « My Digital Identity ». In *ePortfolio & Identité Numérique 2008*, Montréal, 5-7 mai 2008.
- Emin Jean-Claude (2008). « Que fait-on des évaluations internationales dans le système éducatif français ? ». In Baillat Gilles, De Ketele Jean-Marie, Paquay Léopold & Thélot Claude (dir.). *Évaluer pour former. Outils, dispositifs et acteurs.* Bruxelles : De Boeck.
- Flutter Julia & Rudduck Jean (2004). *Consulting Pupils: What's in It for Schools?* London : RoutledgeFalmer.
- Gérard François-Marie (2008). « Les outils d'évaluation ouverts, ou la nécessité de clés de fermeture ». In *Évaluer pour former. Outils, dispositifs et acteurs.* Bruxelles : De Boeck, p. 99-110.
- Gilliéron Giroud Patricia (2007). « Le dossier d'évaluation ou portfolio en Suisse romande : Visées institutionnelles, mises en œuvre et perspectives d'avenir ». In *Actes du 19e colloque de l'ADMEE-Europe*, Université du Luxembourg, 11-13 septembre 2006.
- Gordon George (2005). *Responding to Student Needs: Student Evaluation and Feedback Toolkit*. Scotland : Quality Assurance Agency for Higher Education (QAA).
- Goupil Georgette (1998). *Portfolios et dossiers d'apprentissage*. Montréal : Editions Chenelière éducation.
- Hargreaves Eleanore (2005). « Assessment for learning? Thinking outside the (black) box ». *Cambridge Journal of Education*, vol. 35, n° 2, p. 213-224.
- Hartnell-Young Elizabeth (2007). *Impact Study of e-portfolios on learning*. Coventry : Becta.
- Harvey Lee (2001). *Student Feedback: A report to the Higher Education Funding Council for England*. Birmingham : University of Central England.
- Hoban Garry & Hastings Geoff (2006). « Developing different forms of student feedback to promote teacher reflection: A 10-year collaboration ». *Teaching and Teacher Education*, vol. 22, n° 8, p. 1006-1019.
- Hoskins Bryony & Fredriksson Ulf (2008). *Learning to Learn: What is it and can it be measured?*. Ispra : Centre for Research on Lifelong Learning (CRELL).
- Hoskins Bryony & Deakin Crick Ruth (2008). *Learning to Learn and Civic Competences: Different currencies or two sides of the same coin?* Ispra : Centre for Research on Lifelong Learning (CRELL).
- Houchot Alain & Robine Florence (2007). *Les livrets de compétences : Nouveaux outils pour l'évaluation des acquis*. Paris : Ministère de l'Éducation nationale, n° 2007- 048.
- Klenowski Val (2002). *Developing Portfolios for Learning and Assessment: Processes and Principles*. London : Routledge.
- Leitch Ruth & Odena Oscar (2007). « Consulting secondary school students on increasing participation in their own assessment in Northern Ireland ». In *Annual European Conference on Educational Research*, Ghent, septembre 2007.
- Little David (2003). *The European Language Portfolio in use: nine examples*. Strasbourg : Council of Europe.
- Mangez Éric (2008). *Réformer les contenus d'enseignement*. Paris : Presses universitaires de France.
- McIntyre Donald, Pedder David & Rudduck Jean (2005). « Pupil voice: comfortable and uncomfortable learnings for teachers ». *Research Papers in Education*, vol. 20, n° 2, p. 149-168.

- Merle Pierre (2007). *Les notes. Secrets de fabrication*. Paris : Presses universitaires de France.
- Morlaix Sophie (2007). *Identifier et évaluer les compétences dans le système éducatif : quels apports pour la recherche en education ?* [Rapport d'Habilitation à diriger des recherches]. Dijon : Institut de Recherche sur l'Éducation.
- Nicol David J. & Macfarlane-Dick Debra (2006). « Formative assessment and self-regulated learning: A model and seven principles of good feedback practice ». *Studies in Higher Education*, vol. 31, n° 2, p. 199-218.
- Paquay Léopold, De Ketele Jean-Marie, Thélot Claude & Baillat Gilles (dir.) (2008). *Évaluer pour former*. Bruxelles : De Boeck.
- Prebble Tom, Hargraves Helen & Leach Linda *et al.* (2005). *Impact of Student Support Services and Academic Development Programmes on Student Outcomes in Undergraduate Tertiary Study: A Synthesis of the Research: Report to the Ministry of Education*. New Zealand: Ministry of Education.
- Pryor John & Crossouard Barbara (2008). « A socio-cultural theorisation of formative assessment ». *Oxford Review of Education*, vol. 34, n° 1, p. 1-20.
- Rey Bernard, Carette Vincent, Defrance Anne & Kahn Sabine (2006). *Les compétences à l'école. Apprentissage et évaluation*. Bruxelles : De Boeck.
- Rey Bernard (2008). « Quelques aspects éthiques de l'évaluation ». In Baillat Gilles, De Ketele Jean-Marie, Paquay Léopold & Thélot Claude (dir.). *Évaluer pour former. Outils, dispositifs et acteurs.* Bruxelles : De Boeck, p. 57-67.
- Rey Olivier (2008). « De la transmission des savoirs à l'approche par compétences ». *Dossier d'actualité de la VST*, n° 34. En ligne (consulté le 13 novembre 2008) : <http://www.inrp.fr/vst/LettreVST/34_avril2008.htm>.
- Richardson John T. E. (2005). « Instruments for obtaining student feedback: A review of the literature ». *Assessment & Evaluation in Higher Education*, vol. 30, n° 4, p. 387-415.
- Sadler D. Royce (1989). « Formative assessment and the design of instructional systems ». *Instructional Science*, vol. 18, n° 2, p. 119-144.
- Scallon Gérard (2007). *L'évaluation des apprentissages dans une approche par compétences*. 2e éd. Bruxelles : De Boeck.
- Scallon Gérard (2008). « Évaluation formative et psychologie cognitive : Mouvances et tendances ». In Grégoire Jacques (dir.). *Évaluer les apprentissages. Les apports de la psychologie cognitive*. Bruxelles : De Boeck.
- Schärer Rolf (2007). *Portfolio européen des langues : Projet de rapport intérimaire 2005-2006. Résumé analytique*. Strasbourg : Conseil de l'Europe.
- Scheuermann & Guimarães Pereira (2008). *Towards a Research Agenda on Computer-based Assessment: Challenges and needs for European Educational Measurement*. Ispra: Centre for Research on Lifelong Learning (CRELL).
- Sebba Judy, Deakin Crick Ruth & Yu Guoxing *et al.* (2008). *Systematic review of research evidence of the impact on students in secondary schools of self and peer assessment*. London : EPPI-Centre.
- Siemens George (2004). « EPortfolios ». *Elearnspace*. En ligne (consulté le 19 novembre 2008) : <http://www.elearnspace.org/Articles/eportfolios.htm>.
- Taras Maddalena (2005). « Assessment summative and formative: Some theoretical reflections ». *British Journal of Educational Studies*, vol. 55, n° 4, p. 466-478.
- Tardif Maurice (2006). *L'évaluation des compétences : Documenter le parcours de développement*. Montréal : Chenelière Éducation.
- Weeden Paul & Winter Jan (1999). *Learners' Expectations of Assessment for Learning: Report for the Qualifications and Curriculum Authority*. University of Bristol.
- White Barbara Y. & Frederiksen John R. (1998). « Inquiry, Modeling, and Metacognition: Making Science Accessible to All Students ». *Cognition and Instruction*, vol. 16, n° 1, p. 3-118.
- Wiliam Dylan (2007). « Changing Classroom Practice ». *Educational Leadership*, vol. 65, n° 4, p. 36-42.

Rédacteurs : Laure Endrizzi et Olivier Rey

Ce *Dossier d'actualité* est une publication mensuelle du service de Veille scientifique et technologique de l'Institut national de recherche pédagogique, © INRP.

- Vous abonner ou vous désabonner http://www.inrp.fr/vst/LettreVST/Abonnement.htm
- Nous contacter http://www.inrp.fr/vst/Contact.php?nouveau=oui
- Consulter les nouveautés http://www.inrp.fr/vst/Dernieres_MAJ.php

Veille scientifique et technologique
Institut national de recherche pédagogique
19, allée de Fontenay – BP 17424 – 69347 Lyon cedex 07
Tél. : +33 (0)4 72 76 61 00 – Fax. : +33 (0)4 72 76 61 93