

Chapitre 6

Le rôle de l'évaluation formative dans les environnements d'apprentissage efficaces

Dylan Wiliam

Institute of Education, Université de Londres

Dylan Wiliam décrit l'évaluation comme une passerelle entre l'enseignement et l'apprentissage. Le concept d'évaluation formative est né de la reconnaissance de la valeur du feedback. Il recourt aux métaphores de la navigation, par exemple « maintenir le cap », pour éclairer son propos. Il est amplement démontré, comme nous le verrons dans ce chapitre, que le feedback améliore l'apprentissage ; mais la plupart des études souffrent d'un manque d'analyses théoriques et négligent les conséquences à long terme. La définition ci-dessous met l'accent sur le rôle de l'évaluation dans l'amélioration de la qualité des décisions pédagogiques. Cinq « stratégies clés » entrent en jeu :

- 1. Clarifier, communiquer et comprendre les intentions d'apprentissage et les critères de réussite.*
- 2. Concevoir des activités de classe qui produisent des données sur les acquisitions.*
- 3. Donner un feedback qui aide les apprenants à progresser.*
- 4. Inciter les élèves à devenir des personnes ressources l'un pour l'autre.*
- 5. Encourager les élèves à s'approprier leur apprentissage.*

L'évaluation formative est envisagée comme un processus de mise à profit de « moments de contingence » dans le but de réguler les processus d'apprentissage.

Introduction

L'évaluation a de multiples fonctions dans nos sociétés modernes, notamment la certification des acquis des élèves et l'imposition d'une obligation de résultat à l'institution scolaire. Depuis une quarantaine d'années cependant, on observe aussi un intérêt grandissant pour le rôle qu'elle pourrait jouer dans le soutien à l'apprentissage ; apparaît le terme « évaluation formative » ou « évaluation pour l'apprentissage ». Ce chapitre fait un rapide tour d'horizon de l'évolution récente du concept d'évaluation formative. Il analyse en particulier comment l'idée centrale du feedback s'est élargie aux pratiques en classe. Il présente des données relatives aux effets de l'évaluation formative sur l'apprentissage et examine les problèmes de définition. Il analyse enfin le lien qu'entretient l'évaluation formative avec le design pédagogique à travers la « régulation » des processus d'apprentissage.

L'évaluation au cœur de l'apprentissage

Si l'on pouvait prédire ce que les élèves ont assimilé à la fin d'une séquence, il serait facile de concevoir l'apprentissage. Il suffirait en effet de vérifier qu'ils disposent des connaissances préalables nécessaires à une leçon sur un thème donné pour être sûr que tous auraient acquis les compétences visées une fois exécutées les tâches proposées. Denvir et Brown (1986a ; 1986b) démontrent cependant que même si l'enseignant conçoit des activités d'apprentissage bien adaptées à l'acquisition de compétences particulières, et même s'il prend en considération les acquis des élèves, le résultat observé s'écarte souvent de l'objectif fixé.

Pourtant, la plupart des enseignants ne mesurent les acquis qu'à la fin de la séquence d'apprentissage. Imagine-t-on l'équipage d'un avion long-courrier se concentrer sur le cap à suivre entre l'aéroport de départ et celui de destination sans jamais faire le point en cours de vol ? Tous les pilotes savent bien que cette stratégie n'est pas fiable. C'est pourquoi l'équipage ne se contente pas de définir le meilleur plan de vol, mais procède régulièrement à des relevés de position et ajuste la trajectoire en fonction de la situation.

Dans le même esprit, Benjamin Bloom proposait il y a plus de quarante ans de compléter l'évaluation de fin de séquence destinée à mesurer les acquisitions par une évaluation à chaque étape du processus d'enseignement et d'apprentissage pour apporter un feedback et les correctifs nécessaires (Bloom, 1969 p. 48). Il notait également que si une évaluation de ce type « pouvait éventuellement faire l'objet d'une notation et entrer pour une part dans le processus de jugement et de classement », elle serait infiniment plus efficace « si elle était exclue du processus de notation et utilisée essentiellement comme une aide à l'enseignement » (p. 48).

Il y a fort longtemps, David Ausubel déclarait déjà : « Si je devais réduire la psychologie scolaire à un seul principe, je dirais ceci : le premier facteur d'influence sur l'apprentissage est ce que l'apprenant sait déjà. Alors vérifions ses connaissances et agissons en conséquence » (Ausubel, 1968, p. iv). L'évaluation est cruciale pour un apprentissage efficace, car même si les élèves débutent tous approximativement au même niveau dans une matière, ils se retrouvent très rapidement à des points différents car leurs acquisitions ne sont pas identiques.

L'idée fondamentale que j'analyse dans ce chapitre est la suivante : la conception des situations d'apprentissage doit intégrer l'imprévisibilité des acquisitions. L'évaluation a donc un rôle clé à jouer en établissant un lien entre les activités scolaires programmées par l'enseignant et les gains de compétences qui en résultent chez les élèves. En d'autres termes, l'évaluation sert de passerelle entre l'enseignement et l'apprentissage. Mon objectif est ici de proposer une base théorique claire expliquant l'aide que l'évaluation peut apporter à l'apprentissage, de montrer comment les différentes formulations de la notion d'évaluation formative proposées au cours des quarante dernières années peuvent s'intégrer dans un cadre conceptuel plus large et d'utiliser ce cadre pour appréhender la recherche dans les domaines voisins.

L'évaluation formative en tant que feedback

La correction de trajectoire évoquée plus haut en matière de navigation aérienne est un exemple de « feedback », un système développé à l'origine dans le domaine de l'ingénierie des systèmes (voir Wiener, 1948). Wiener constate que le « feedback » (rétroaction) conforte parfois un système dans la trajectoire qu'il a déjà empruntée, par exemple croissance d'une population qui vit dans l'abondance de nourriture et sans prédateur ou spirale inflationniste des prix et des salaires en économie. Ce feedback est qualifié de « positif » car son effet et la tendance du système sont orientés dans le même sens. Dans d'autres situations, l'effet du feedback s'oppose à la tendance, ce qui a pour conséquence de rétablir la stabilité du système : impact des restrictions de nourriture sur la croissance de la population ou du thermostat d'ambiance dans une pièce. Les ingénieurs qualifient ce feedback de « négatif » car l'effet évolue dans le sens opposé à la tendance du système. En ingénierie, le feedback positif est stérile car il est synonyme d'instabilité et donc de croissance explosive ou d'effondrement. En revanche, le feedback négatif contribue à ramener le système à l'équilibre.

La métaphore du « feedback » est largement convoquée dans le monde de l'éducation. Il convient toutefois de noter que ce terme est employé très différemment en ingénierie et dans l'enseignement. Premièrement, il y a feedback pour un ingénieur lorsque le système peut utiliser les informations pour

influer sur son fonctionnement. Le feedback est « une information sur l'écart entre le niveau atteint et le niveau de référence d'un paramètre du système, que l'on utilise pour réduire cet écart (Ramaprasad, 1983, p. 4). En revanche, les enseignants emploient souvent le terme « feedback » pour décrire toute information fournie à l'apprenant sur ses performances, que cette information permette ou non de réduire l'écart (Sadler, 1989). Autrement dit, si nous prenons le terme au sens de l'ingénieur, le feedback n'est pas une simple information donnée à l'apprenant sur ses performances, il doit orienter les actions de ce dernier de manière productive.

Deuxièmement, les qualificatifs « positif » et « négatif » ont eux aussi un sens légèrement différent. En ingénierie, on se réfère à l'effet du feedback par rapport à la tendance du système. Dans le domaine de l'éducation, ces termes tendent à être employés comme des jugements de valeur sur les effets du feedback. Le feedback qui indique que l'apprenant est sur la bonne voie et qui vient ainsi renforcer l'apprentissage sera qualifié de « positif », tant par les éducateurs que par les ingénieurs. Toutefois, prenons la situation d'un élève qui relâche ses efforts après avoir reçu des appréciations critiques, obtient des appréciations plus mauvaises encore et finit par se démobiliser complètement. Pour un éducateur, il s'agit d'un exemple de feedback négatif, mais pour un ingénieur, le feedback est au contraire positif car il entraîne le système (l'élève) dans la même direction que lui.

Troisièmement, et c'est sans doute le plus important, nous voulons, en éducation, encourager l'autonomie de l'apprentissage – pour que les élèves développent leurs compétences d'autorégulation et aient ainsi moins besoin de feedback. En revanche, personne ne viendrait contester l'utilité d'un thermostat d'ambiance parce que la chaudière n'a pas appris à déterminer elle-même quand se mettre en route ou s'arrêter.

Ces distinctions pourraient nous sembler purement sémantiques, mais elles se situent en fait au cœur des problèmes rencontrés dans l'élaboration de systèmes efficaces de feedback dans l'enseignement. Crooks (1988), dans une méta-analyse portant sur plus de deux cents études de terrain, a examiné l'impact des pratiques d'évaluation sur les élèves. Il en conclut que le potentiel de guidage de l'apprentissage inhérent à l'évaluation n'est pas réalisé parce que la fonction sommative de l'évaluation, qui vise à attribuer des notes et à mesurer le volume des connaissances acquises, reste prédominante.

Données sur les effets du feedback

Les études montrent que le feedback peut sensiblement améliorer les résultats scolaires, mais un certain nombre de réserves s'imposent en guise d'introduction. Dans la plupart des études, les résultats présentés correspondent à « l'ampleur de l'effet normalisé » (ou « ampleur de l'effet ») : selon

Cohen (1988), elle correspond à la valeur de l'écart des performances entre deux groupes (ceux qui reçoivent un feedback et ceux qui n'en reçoivent pas par exemple) divisée par une mesure de la dispersion des scores de la population (écart-type). Si l'ampleur de l'effet normalisé présente des avantages indéniables par rapport à la signification statistique atteinte dans les études comparatives (Harlow, Mulaik et Steiger, 1997), ce système de mesure souffre de quelques lacunes lorsqu'il s'agit de comparer les résultats d'études expérimentales différentes. En effet, en particulier lorsque la cohorte est limitée (études portant sur certaines sous-populations, par exemple élèves ayant des besoins éducatifs spéciaux), l'ampleur de l'effet est artificiellement gonflée car le diviseur utilisé dans le calcul est plus petit (Black et Wiliam, 1998a). D'autre part, les mesures des acquisitions diffèrent beaucoup en termes de sensibilité aux effets de l'enseignement, ou selon que la mesure porte directement sur ce que les élèves ont appris ou s'en éloigne, comme c'est le cas dans la plupart des tests et examens nationaux (Wiliam, 2008). D'où la difficulté de proposer des règles strictes quant à l'interprétation des effets observés. De manière générale toutefois, du moins pour les mesures normalisées du rendement scolaire, une valeur d'effet voisine de 0,4, valeur type dans les études sur le feedback, indique une augmentation d'au moins 50 % du rythme d'apprentissage. En d'autres termes, les élèves apprennent en huit mois ce que d'autres mettent un an à assimiler. Il s'agit donc d'une augmentation substantielle de la productivité éducative, surtout si elle peut être observée à l'échelle nationale.

Une réserve d'ordre plus général tient à la multiplicité des objectifs de l'évaluation scolaire. En effet, les études comparatives peuvent aboutir à des conclusions trompeuses si la comparaison porte sur des fonctions pour lesquelles l'évaluation n'était pas conçue (voir par exemple Natriello, 1987). Ainsi, constater qu'un feedback différencié a plus d'impact sur l'orientation de l'apprentissage futur de l'apprenant que sur ses notes limite sans doute la démonstration à la plus grande efficacité des systèmes lorsqu'ils sont utilisés aux fins pour lesquelles ils ont été prévus.

Nonobstant ces restrictions, le premier constat important est qu'une évaluation régulière peut avoir une forte incidence sur l'apprentissage. Ainsi, les élèves soumis à un contrôle au moins toutes les quinze semaines enregistrent un score supérieur de 0,5 écart type à celui des élèves qui n'ont effectué aucun contrôle ; et si on observe que ces contrôles fréquents sont associés à des niveaux plus élevés de réussite, une périodicité des contrôles inférieure à deux semaines n'apporte aucun bénéfice supplémentaire (Bangert Drowns, Kulik, Kulik et Morgan, 1991). La qualité du feedback et son utilisation constituent des paramètres beaucoup plus importants que la fréquence. L'analyse d'une quarantaine de rapports de recherche sur les effets du feedback dans des évaluations de type contrôle (questions insérées dans le matériel d'apprentissage programmé ou contrôles récapitulatifs à la fin d'un

module d'apprentissage) montrent en effet que le type et les modalités de communication du feedback sont deux éléments cruciaux (Bangert-Drowns *et al.*, 1991). Lorsque les élèves pouvaient anticiper et consulter furtivement les réponses avant de chercher à répondre aux questions, leurs acquisitions étaient inférieures à ce qu'elles auraient été si les études avaient pris en compte cette « possibilité de recherche préalable » (valeur de l'effet : 0,26). Mais surtout, on constate que lorsque le feedback propose une analyse détaillée de la bonne réponse, les élèves apprennent plus que lorsqu'on se contente de leur indiquer si la réponse est juste ou fausse (valeur de l'effet : 0,58).

Le feedback peut aussi s'avérer utile pour les enseignants. Fuchs et Fuchs (1986) ont réalisé une méta-analyse portant sur 21 rapports concernant le recours au feedback donné aux enseignants et par eux, à raison de deux à cinq fois par semaine. L'effet moyen sur les résultats entre les groupes étudiés et les groupes témoins est de 0,70 écart-type. Dans la moitié des études analysées environ, les enseignants avaient fixé les règles à respecter concernant l'analyse des données et les actions à entreprendre ; l'effet moyen était alors significativement supérieur et atteignait 0,92. Lorsque les actions avaient été laissées à la libre appréciation de l'enseignant, l'effet n'excédait pas 0,42. Dans les études où les enseignants produisaient des graphiques de progression pour chaque élève afin de les orienter et de les encourager, l'effet était supérieur (effet moyen égal à 0,70) à celui relevé dans les études où les enseignants n'avaient pas procédé ainsi (effet moyen égal à 0,26).

Ces résultats semblent affectés par le type d'apprentissage étudié. Dempster (1991) constate que de nombreuses études mesurent les acquis en termes de connaissance déclaratives et de compétences élémentaires. Il n'est donc pas certain que ces conclusions puissent être étendues aux compétences de réflexion d'ordre supérieur. Dans une autre communication (Dempster, 1992), il fera valoir que l'évaluation n'est pas encore suffisamment prise en compte dans l'enseignement en dépit des avantages évidents qu'elle représente, et ce malgré le consensus qui commence à se faire jour au sein de la communauté scientifique sur les conditions de mise en place d'une évaluation efficace (contrôles fréquents peu après le cours, exigences accrues, feedback donné peu après les contrôles). Il constate que les pratiques pédagogiques actuelles sont loin d'avoir atteint cet idéal.

Une méta-analyse d'Elshout-Mohr (1994), initialement publiée en néerlandais et regroupant nombre d'études non disponibles en anglais, semble indiquer que pour les tâches plus complexes, connaître la bonne réponse n'est pas aussi utile que pour les tâches simples. L'apprentissage ne se borne pas à corriger les erreurs mais aussi à développer de nouvelles compétences. Le feedback doit donc se présenter sous forme de dialogue et ne pas se limiter à donner la réponse juste. L'apprenant devient donc un élément actif du processus d'apprentissage.

Une part importante de ces travaux s'est attachée aux effets du feedback à l'école. En 1996, Kluger et DeNisi ont publié une étude consacrée au feedback à l'école, à l'université et au travail¹. Dans l'ensemble, l'effet moyen est de 0,41 écart-type, mais cette valeur varie considérablement d'une étude à l'autre. Et pour 50 des 131 études prises en compte (38 %), le feedback a en réalité **abaissé** la performance moyenne.

Dans le cadre d'un programme de recherches plus vaste visant à développer des environnements de tutorat intelligents, Shute (2008) a examiné les résultats des recherches portant sur le feedback donné aux élèves². Ce tour d'horizon a permis de relever les principales divergences dans la littérature et, comme on pouvait s'y attendre, conclut qu'il n'existe pas de réponse simple à la question suivante : « Quel type de feedback fonctionne ? ». Il a cependant repris à son compte les résultats d'études antérieures sur l'ampleur de l'effet qu'on peut attendre du feedback (effet normalisé de l'ordre de 0,4 à 0,8 écart-type).

Quelques indications pour un feedback efficace

Kluger et DeNisi (1996) ont cherché à comprendre les raisons de la baisse de performance parfois due au feedback et se sont efforcés de définir quels éléments venaient en « modérer » les effets. Ils constatent que le feedback est moins efficace lorsqu'il porte sur l'égo, plus efficace lorsqu'il est dirigé sur la tâche en cours, et encore plus efficace lorsqu'il couvre les détails de la tâche et fixe des objectifs.

Cependant, même les quelques bénéfices relevés par Kluger et DeNisi peuvent parfois s'avérer contre-productifs. Ces auteurs soulignent que le feedback peut pousser l'apprenant à travailler plus, ce qui pourrait sans doute être bénéfique, mais pourrait tout aussi bien le conduire à canaliser ses efforts dans une direction précise, à modifier ou rejeter l'objectif, voire à totalement ignorer le feedback. Et même lorsque le feedback a un effet positif sur l'apprentissage, ce peut être en insistant sur les objectifs utilitaires au détriment d'un apprentissage en profondeur. Dans leur conclusion, ils avancent qu'il est plus important d'analyser les processus induits par le feedback que de chercher à établir si le feedback en général améliore les performances.

Shute (2008) propose quelques « principes directeurs liminaires » destinés à favoriser l'apprentissage et en déterminer le moment propice pour la conception d'un feedback efficace.

Le feedback ne doit pas se concentrer sur l'apprenant mais s'intéresser à tous les éléments spécifiques d'une tâche et proposer des pistes d'amélioration ; il doit mettre en avant le « quoi, comment et pourquoi » d'un problème et ne pas se contenter d'indiquer si le résultat est correct ou non. Un feedback approfondi doit se présenter sous forme d'unités facilement maîtrisables, et pour paraphraser la célèbre maxime d'Einstein, il doit être « aussi simple que possible mais pas plus simple ». Toutefois, il ne doit pas être trop détaillé ni

trop précis pour ne pas « étayer » l'apprentissage au point d'éviter à l'apprenant de penser par lui-même. En outre, il est plus efficace lorsqu'il est dispensé par une source en laquelle l'apprenant a confiance (humaine ou informatique).

Le moment optimal dépend beaucoup du type d'apprentissage en jeu : un feedback immédiat est particulièrement utile pour l'apprentissage de procédures ou lorsque la tâche dépasse de loin les capacités initiales de l'apprenant. Le feedback différé est en revanche plus adapté aux tâches à la portée de l'apprenant ou en cas de transposition à d'autres contextes.

Dans une étude récente, Hattie et Timperley (2007) considèrent que l'objectif du feedback consiste à réduire l'écart entre les acquis ou les performances et l'objectif visé (comme le propose Ramaprasad, 1983). S'appuyant sur les travaux de Deci et Ryan (1994) et Kluger et DeNisi (1996), leur modèle postule que les élèves peuvent réduire l'écart soit en mettant en œuvre des stratégies plus efficaces ou en redoublant d'efforts, soit en abandonnant, en estompant ou en réduisant les objectifs qu'ils s'étaient fixés. L'enseignant peut diminuer l'écart en modifiant la difficulté ou la spécificité du but, ou en apportant un plus grand soutien aux élèves. Ce modèle précise trois types de questions auxquelles le feedback est censé répondre (Où vais-je ? Comment ? Et après ?). Chaque question posée fonctionne à quatre niveaux : feedback sur la tâche, feedback sur le traitement de la tâche, feedback sur l'autorégulation et feedback sur le soi en tant que personne. Ils démontrent que le feedback sur le soi est la forme la moins efficace de feedback ; que les feedbacks qui portent sur l'autorégulation et sur le traitement de la tâche sont performants en termes de traitement approfondi et de maîtrise des tâches et que le feedback sur la tâche est efficace s'il vise à améliorer le traitement des stratégies ou renforce l'autorégulation (reste que ces conditions sont rarement réunies dans la pratique).

L'évaluation formative, partie intégrante de l'enseignement

La revue présentée ci-dessus montre que certaines formes de feedback donné aux apprenants en cours d'apprentissage ont des effets positifs, mais que ceux-ci ne peuvent pas être tenus pour acquis. Les effets ne dépendent pas uniquement de la qualité du feedback, mais aussi de l'environnement d'apprentissage dans lequel il est donné, des attitudes et des motivations de l'apprenant et de tout un éventail d'autres facteurs contextuels (Boekaerts, dans ce volume). C'est pourquoi lorsque Paul Black et moi-même avons cherché à actualiser les études de Natriello et de Crooks, nous avons résolument choisi d'élargir le champ des recherches (selon nos observations, les études de Natriello de Crooks comportaient respectivement 91 et 241 références, et seules neuf d'entre elles étaient communes aux deux documents, aucune ne mentionnait l'étude conduite par Fuchs et Fuchs). Au lieu de recourir aux méthodes de recherche électronique, nous avons consulté chaque numéro des

76 revues spécialisées susceptibles de contenir des études dans ce domaine parues entre 1987 et 1997. Notre méta-analyse (Black et Wiliam, 1998a), portant sur 250 études, a montré qu'une mise en œuvre efficace de l'évaluation en classe permet une nette amélioration des résultats des élèves, de l'ordre de 0,4 à 0,7 écart-type. Nous avons toutefois observé les problèmes d'interprétation des effets évoqués plus haut.

Black et Wiliam présentent plusieurs « exemples probants » (la méta-analyse de Fuchs et Fuchs et sept études en contexte de classe) qui illustrent les caractéristiques d'une évaluation formative efficace. La plus importante est sans doute que l'évaluation formative doit être intégrée à la pratique de classe, ce qui exige une réorganisation fondamentale des activités pédagogiques :

Il paraît difficile d'innover en matière d'évaluation formative sans un changement profond du travail en classe. Tout travail exige un certain degré de feedback entre enseignés et enseignant, ce qui signifie que la qualité de leurs interactions doit être placée au centre de la pédagogie. (Black et Wiliam, 1998a, p. 16)

Nous avons en outre relevé que pour que l'évaluation soit formative, il faut que les données du feedback soient exploitées ; les traitements différenciés mis en place suite au feedback sont donc au cœur d'un apprentissage efficace. De plus, les enseignants ont besoin de modèles adaptés qui anticipent les éventuelles réactions des élèves au feedback et la manière dont ils s'en servent, pour pouvoir sélectionner les traitements différenciés les plus appropriés à la situation de chacun. Comme le remarque Perrenoud (1998) dans son commentaire sur l'article de Black et Wiliam : « une partie des feedbacks donnés aux élèves en classe sont comme des bouteilles lancées à la mer. Nul n'a la certitude que le message qu'elles contiennent trouvera un jour un destinataire. »

Pour résoudre ce problème, nous avons examiné la perspective de l'élève, le rôle de l'enseignant et un certain nombre de systèmes d'organisation de l'enseignement dont l'évaluation formative est une composante majeure. En définissant les implications en termes de politique et de pratique d'évaluation formative, nous concluons :

aucun modèle optimal ne ressort réellement de la présente étude sur lequel [...] éventuellement fonder une politique. Ce qui ressort en revanche, c'est une série de principes directeurs, étant précisé toutefois que loin d'être marginaux, les changements à apporter dans la pratique pédagogique sont fondamentaux et devront être incorporés dans la pratique de chaque enseignant selon des modalités qui lui sont propres [...]. Cela signifie que la réforme dans ce domaine prendra inévitablement du temps et devra recevoir le soutien assidu des praticiens comme des chercheurs (p. 62).

La plupart des travaux analysés par Natriello, Crooks, Kulik et collègues, et Black et Wiliam portent sur des élèves d'âge scolaire (jusqu'à 18 ans). Nyquist (2003) s'est intéressé aux études consacrées aux apprenants de l'enseignement supérieur. Il a passé en revue environ 3 000 études sur les effets du feedback, parmi lesquelles 86 répondaient aux critères fixés, à savoir :

- Manipulation expérimentale d'une caractéristique pertinente en termes de feedback ;
- Échantillon d'apprenants de l'enseignement supérieur ;
- Mesure des performances académiques ;
- Volume suffisant de données quantitatives pour pouvoir calculer l'ampleur de l'effet.

À partir de ces 86 études, il a été possible de calculer 185 effets. Après un certain nombre d'ajustements techniques (visant à limiter à deux écarts-types les valeurs extrêmes par rapport à l'effet moyen et à corriger les biais liés aux échantillons de petites tailles entre les différentes études), l'analyse a produit un effet moyen de 0,40 écart-type, résultat pratiquement identique à celui obtenu par Kluger et DeNisi. Cet effet moyen diminue légèrement et passe à 0,35 (erreur-type = 0,17) après ajustement (pondération de la part respective des effets dans l'effet moyen en proportion de leur fiabilité), même si les effets eux-mêmes varient fortement (écart-type entre - 0,6 et 1,6).

Nyquist a établi la typologie suivante de l'évaluation formative afin d'analyser les éléments « modérateurs » de l'effet :

- **Feedback faible seul** : les élèves sont simplement informés de leur score ou de leur note, souvent qualifié de « connaissance des résultats » ;
- **Feedback seul** : les élèves sont informés de leur score ou de leur note, des objectifs clairs de travail leur sont fixés, ou un feedback leur est apporté sur les réponses aux questions auxquelles ils ont tenté de répondre, souvent qualifié de « connaissance des résultats justes » ;
- **Évaluation formative minimale** : les élèves sont informés des résultats corrects, et quelques explications leur sont données ;
- **Évaluation formative modérée** : les élèves sont informés des résultats corrects, quelques explications leur sont données et quelques suggestions d'amélioration leur sont proposées ;
- **Évaluation formative poussée** : les élèves sont informés des résultats corrects, quelques explications leur sont données, et des activités spécifiques leur sont proposées pour leur permettre de s'améliorer.

L'effet moyen normalisé pour chaque type d'intervention est présenté au tableau 6.1 ci-dessous.

Tableau 6.1. Effets des différents types de feedback

| | Nombre | Effet |
|-------------------------------|--------|-------|
| Feedback faible seul | 31 | 0,14 |
| Feedback seul | 48 | 0,36 |
| Évaluation formative minimale | 49 | 0,26 |
| Évaluation formative modérée | 41 | 0,39 |
| Évaluation formative poussée | 16 | 0,56 |
| Total | 185 | |

Source : Nyquist, 2003. Ces chiffres sont les valeurs corrigées présentées dans une communication personnelle, et non ceux qui étaient indiqués dans la thèse originale.

Les résultats obtenus par Nyquist corroborent les conclusions de Bangert-Drowns *et al.* analysées plus haut. Un feedback qui se borne à dresser l'état des lieux des acquis ne présente pas vraiment d'intérêt, alors que celui qui incite les élèves à s'impliquer dans une activité consciente peut avoir de puissants effets sur l'apprentissage.

Les revues de la recherche conduites par Natriello (1987), Crooks (1988), Bangert-Drowns *et al.* (1991), et Black et Wiliam (1998a) soulignent clairement que les différentes formes de feedback apportées aux élèves ne sont pas toutes aussi efficaces. Prenons un autre exemple : Meisels, Atkins-Burnett, Xue, Bickel et Son (2003) ont étudié l'impact du système d'examen des échantillons de travail (*Work Sample System* – système d'évaluation des performances intégrée au curriculum). Les résultats des élèves sont meilleurs en lecture, mais on n'observe aucune différence significative en mathématiques. Les détails et le mode de mise en œuvre du système utilisé, ainsi que la nature du feedback donné aux apprenants s'avèrent être des variables cruciales, la moindre modification produisant souvent des effets importants sur l'efficacité.

La plupart des études couvertes par les différentes revues se concentrent sur les élèves plus âgés, mais les attitudes vis-à-vis de l'apprentissage sont modelées par le feedback que ces derniers ont reçu dès le plus jeune âge. Dans une étude conduite pendant un an en Angleterre dans huit classes de maternelle et de première année de primaire, Tunstall et Gipps (1996a ; 1996b) ont pointé les différents rôles que joue le feedback. Comme Torrance

et Pryor (1998), ils ont constaté qu'une part importante du feedback apporté aux élèves touchait à la socialisation : « je n'aiderai que ceux qui sont assis et qui lèvent la main » (p. 135). Au-delà de ce rôle de socialisation, ils ont dégagé quatre types de feedback lié au travail scolaire.

Les deux premiers revêtent un caractère essentiellement **évaluatif**. Dans le premier cas, le feedback consiste à distribuer des récompenses et des punitions (par exemple, les élèves sont autorisés à partir plus tôt déjeuner à la cantine s'ils ont bien travaillé, alors qu'ils sont menacés de devoir rester tant qu'ils n'auront pas terminé leur travail). Dans le deuxième, l'enseignant indique un niveau d'appréciation du type « je suis très content de toi aujourd'hui », ou « tu m'as beaucoup déçu aujourd'hui ». Les deux autres types d'évaluation sont de type « **descriptif** ». L'un est centré sur l'adéquation du travail fourni aux critères de réussite fixés par l'enseignant, allant de l'appréciation d'un travail qui remplit tous les critères (« c'est extrêmement bien expliqué ») aux mesures que l'élève doit engager pour s'améliorer (« je veux que tu revoies toutes les opérations et que tu écrives le signe égal partout »). Le dernier mode de feedback met l'accent sur le processus et l'enseignant joue le rôle de facilitateur plutôt que d'évaluateur. Comme l'expliquent Tunstall et Gipps (1996a), les enseignants qui pratiquent ce type de feedback « font prendre conscience du travail en cours, renforcent la perception de ce qui est entrepris et incitent à la réflexion » (p. 399).

La plupart des recherches évoquées plus haut ont été publiées en anglais. Pour que le tour d'horizon soit complet dans ce domaine, l'OCDE a commandé une étude bibliographique des recherches publiées en français (Allal et Lopez, 2005) et en allemand (Köller, 2005) dans le cadre de l'étude sur l'évaluation formative (Looney, 2005).

Allal et Lopez précisent que la recherche en France et dans les régions francophones de Belgique, du Canada et de Suisse s'appuie essentiellement sur des travaux théoriques, auxquels viennent s'ajouter quelques études empiriques contrôlées. La principale conclusion de leur analyse, qui a porté sur 100 études réalisées au cours des trente années précédentes, est que les études sur les pratiques d'évaluation dans les classes francophones font appel à une « conception élargie de l'évaluation formative », dans le droit fil de Black et Wiliam. Allal et Lopez soulignent que l'élément central du feedback dans la tradition anglophone (illustrée par Bloom) est la « remédiation », qu'elles résument en « feedback + correction ». En revanche, dans la plupart des recherches conduites dans les pays francophones, le concept central est la « régulation », résumée en « feedback + adaptation » (p. 269)³.

Allal et Lopez repèrent quatre grands courants dans la littérature scientifique francophone. Le premier, qu'elles qualifient d'orienté sur « les instruments », s'intéresse plus particulièrement à l'élaboration d'outils d'évaluation tels que les banques d'items diagnostics et les systèmes de tests adaptatifs.

Le deuxième, « recherche de cadres théoriques », insiste sur « la recherche de théories susceptibles d'offrir une orientation conceptuelle pour conduire l'évaluation ». Le troisième courant, « études des pratiques existantes dans leurs contextes », vise à donner un ancrage empirique à la recherche de cadres théoriques en l'articulant avec l'étude des pratiques d'évaluation formative en situation de classe. Le quatrième courant, le plus récent, postule « l'engagement actif des élèves dans l'évaluation », à savoir l'auto-évaluation, l'évaluation par les pairs et la construction conjointe de l'évaluation par les élèves et les enseignants.

La notion d'évaluation formative située au cœur de la régulation des processus d'apprentissage a été reprise par certains chercheurs anglophones (voir par exemple Wiliam, 2007), et l'élargissement du concept d'évaluation formative a été relevé par Brookhart (2007). Dans son étude bibliographique sur « l'évaluation formative en situation de classe », elle retrace l'évolution du concept d'évaluation formative comme une série de formulations étroitement imbriquées (p. 44) :

- Information sur le processus d'apprentissage ;
- Information sur le processus d'apprentissage dont l'enseignant peut se servir pour guider ses décisions pédagogiques ;
- Information sur le processus d'apprentissage dont l'enseignant peut se servir pour guider ses décisions pédagogiques et que les élèves peuvent utiliser pour améliorer leurs performances ;
- Information sur le processus d'apprentissage dont l'enseignant peut se servir pour guider ses décisions pédagogiques et que les élèves peuvent utiliser pour améliorer leurs performances selon des modalités propres à les motiver.

Cependant, il ne semble guère y avoir de liens en général entre les travaux très théoriques de la tradition francophone et les études expérimentales, en particulier celles conduites aux États-Unis. Allal et Lopez concluent que les travaux des chercheurs francophones sur l'évaluation formative nécessiteraient un ancrage empirique beaucoup plus solide (p. 282).

La revue de la littérature en langue allemande réalisée par Köller (2005) adopte dans un premier temps la même démarche que celle de Black et Wiliam, à savoir des recherches dans les bases de données en ligne auxquelles s'ajoute un dépouillement minutieux des six plus grandes revues en langue allemande dans ce domaine publiées sur la période 1980-2003. Köller constate que les revues scientifiques traitent des nombreuses avancées dans le champ de l'évaluation formative, mais s'intéressent peu aux effets des pratiques d'évaluation formative sur les élèves, même si quelques constats de la littérature en langue anglaise sont corroborés. Il fait état des travaux de

Meyer qui, comme Kluger et DeNisi, constate que les félicitations peuvent nuire à l'apprentissage, alors que les critiques, voire le blâme, peuvent parfois être utiles. Un autre volet important analysé par Köller porte sur les différentes « normes de référence » adoptées par les enseignants. Plusieurs études, en particulier les travaux de Rheinberg, montrent que les élèves progressent plus avec un enseignant qui les juge en fonction de leurs résultats antérieurs (norme de référence individuelle) qu'avec un enseignant qui compare leurs résultats avec ceux des autres élèves de la classe (norme de référence sociale).

Synthèses théoriques : évaluation formative et évaluation pour l'apprentissage

Depuis une douzaine d'années environ, plusieurs définitions du terme « évaluation formative » ont été proposées. Selon Black et Wiliam (1998a), l'évaluation formative « englobe l'ensemble des activités entreprises par les enseignants ou les élèves qui donnent des informations susceptibles d'être utilisées comme feedback pour modifier les activités d'enseignement et d'apprentissage dans lesquelles ils sont engagés » (p. 7). Cowie et Bell (1999) proposent une définition légèrement plus restrictive et limitent le terme à une évaluation conduite et **suivie de mesures concrètes** en cours d'apprentissage et définissent l'évaluation formative comme étant le « processus mis en œuvre par les enseignants et les élèves pour constater les acquisitions et y réagir afin d'améliorer l'apprentissage **au cours de l'apprentissage** » (p. 32, c'est moi qui souligne). Shepard, Hammerness, Darling-Hammond, Rust, Snowden, Gordon, Gutierrez et Pacheco (2005) adoptent la même vision de l'évaluation formative, à savoir « une évaluation réalisée au cours du processus d'enseignement aux fins d'améliorer l'enseignement ou l'apprentissage » (p. 275).

L'étude de l'OCDE consacrée aux pratiques de l'évaluation formative dans huit systèmes nationaux et provinciaux souligne elle aussi que l'évaluation doit intervenir pendant le cours : « L'évaluation formative désigne les évaluations interactives fréquentes des progrès et des acquis des élèves dans le but d'identifier les besoins d'apprentissage et d'ajuster l'enseignement en conséquence » (Looney, 2005, p. 21). Dans le même esprit, Kahl (2005) écrit : « l'évaluation formative est un outil à la disposition des enseignants pour mesurer la maîtrise qu'ont les élèves des matières et compétences qui leur sont enseignées. Il s'agit d'un outil d'évaluation à mi-parcours visant à repérer les confusions et les erreurs pendant le cours » (p. 11).

Broadfoot, Daugherty, Gardner, Gipps, Harlen, James et Stobart (1999) affirment que pour améliorer les apprentissages, l'évaluation formative doit intégrer cinq facteurs clés : (1) feedback efficace sur les performances de l'élève; (2) implication active des élèves dans leur apprentissage; (3) adaptation de l'enseignement en fonction des résultats de l'évaluation; (4) reconnaissance

de la profonde influence que l'évaluation peut avoir sur la motivation et l'estime de soi des élèves, les deux ayant une influence cruciale sur l'apprentissage et (5) nécessité pour les élèves de s'évaluer eux-mêmes et de comprendre comment ils peuvent progresser. Ils considèrent que le terme « évaluation formative » ne décrit pas ces différents usages de l'évaluation car l'adjectif « formative » laisse le champ à tout un éventail d'interprétations et indique seulement que les évaluations sont fréquentes et programmées en même temps que les cours (p. 7). Ils proposent de le remplacer par le terme « évaluation pour l'apprentissage », comme James l'avait initialement proposé (1992).

Black, Harrison, Lee, Marshall et Wiliam (2004) suggèrent de conserver les deux termes en ce que « l'évaluation pour l'apprentissage » renvoie à toute évaluation dont la priorité, tant d'un point de vue théorique que pratique, vise à promouvoir l'apprentissage des élèves et qu'elle devient « évaluation formative » lorsque les données recueillies sont véritablement utilisées pour adapter l'enseignement aux besoins des élèves (p. S10).

Cela posé, je propose, à la suite de Black et Wiliam (2009), une nouvelle définition qui regrouperait les définitions précédentes et en élargirait le champ : « l'évaluation est formative dans la mesure où les informations qu'elle donne sur les acquisitions des élèves sont suscitées, interprétées et utilisées par les enseignants, les élèves ou leurs pairs pour prendre des décisions sur les étapes suivantes susceptibles d'être meilleures ou mieux étayées que celles qu'ils auraient prises en l'absence de ces données ».

Certains aspects de cette définition méritent qu'on s'y arrête :

- Elle se fonde sur le rôle que jouent les informations dégagées par l'évaluation, plutôt que sur une propriété de l'évaluation elle-même ;
- L'évaluation peut être conduite par l'enseignant, l'élève ou ses pairs ;
- La définition insiste sur les décisions relatives aux étapes suivantes de l'enseignement, plutôt que sur les intentions ou les résultats ;
- La définition est de nature probabiliste ;
- L'évaluation ne modifie pas nécessairement l'orientation de l'enseignement (elle peut simplement confirmer que les actions programmées pour la suite sont adaptées).

Toute évaluation qui produit des données susceptibles d'améliorer les décisions prises par les enseignants, les élèves ou leurs pairs peut donc être formative. Prenons une classe qui a passé un contrôle visant à évaluer la capacité de trouver la fraction la plus grande ou la plus petite dans une série de fractions. Les scores bruts des élèves fourniraient une « évaluation de contrôle », indiquant quels élèves auraient besoin d'un cours ou d'explications complémentaires. Si l'enseignant remarque en outre que de nombreux élèves

dont les scores sont faibles ont mieux réussi les exercices avec des fractions unitaires (le numérateur est égal à 1) qu'avec les fractions plus complexes, il s'agit d'une évaluation diagnostique qui donne des informations particulières sur l'origine des difficultés. L'enseignant peut alors concentrer ses explications complémentaires sur les fractions non unitaires. Si d'après les réponses données, il constate que de nombreux élèves pensent que la fraction la plus petite est celle qui a le plus grand dénominateur et qu'à l'inverse, la plus grande est celle qui a le plus petit dénominateur – stratégie qui fonctionne avec les fractions unitaires (Vinner, 1997), il dispose d'informations « plus malléables à l'intervention pédagogique ». Ces évaluations et leur interprétation ne se bornent pas à signaler le problème (contrôle) et à le localiser (diagnostic), mais elles le situent aussi dans une théorie d'action qui suggère les mesures à prendre pour améliorer l'apprentissage. Les meilleures évaluations formatives sont donc plus prospectives que rétrospectives en ce qu'elles dégagent des solutions pour les actions à venir.

Toute évaluation peut donc être formative dans la mesure où elle vient en appui de décisions qui n'auraient pu être prises ou qui n'auraient pas produit d'aussi bons résultats sans les informations ainsi recueillies. Il ne s'ensuit pas toutefois que l'utilisation à des fins formatives des données recueillies par l'évaluation présente toujours une efficacité constante. Par définition, les évaluations à visée diagnostique peuvent produire de meilleures décisions en termes d'enseignement que celles qui contrôlent simplement les progrès des élèves, et celles qui donnent des indications malléables à l'intervention pédagogique sont, selon toute vraisemblance, encore meilleures.

Les évaluations qui contrôlent, diagnostiquent ou guident se différencient notamment par la spécificité des informations obtenues : pour être malléables à l'intervention pédagogique, l'évaluation ne doit pas se borner à constater qu'il y a ou non acquisition, et dans ce dernier cas, à déterminer ce qui n'a pas été assimilé, elle doit aussi tenir compte à la fois des théories du curriculum et de l'apprentissage. En effet, l'enjeu est bien de répondre à la question « et ensuite ? », ce qui suppose d'avoir une idée claire de la progression de l'apprentissage : une description des « connaissances, compétences, représentations, attitudes ou valeurs que les élèves acquièrent dans une discipline et l'ordre dans lequel ils les acquièrent généralement » (Forster et Masters, 2004, p. 65). La malléabilité à l'intervention pédagogique nécessite aussi une théorie de l'apprentissage car, pour prendre une décision quant au choix des données à recueillir, il ne faut pas s'arrêter à ce qui vient après dans les apprentissages, il faut aussi envisager quels types de difficultés les élèves auront dans les phases suivantes. On trouvera une analyse plus détaillée des liens entre l'évaluation formative et les théories de l'apprentissage dans Black et Wiliam (2005), Brookhart (2007), Wiliam (2007), et Black et Wiliam (2009).

Longueur des cycles d'évaluation formative

Dans l'exemple du contrôle sur les fractions analysé plus haut, l'intervention de l'enseignant suit rapidement le recueil des données sur les performances des élèves. Toutefois, la définition de l'évaluation formative proposée ci-dessus ne fixe pas de durée précise concernant les cycles de recueil de données, d'interprétation et d'action ; la seule exigence est que les informations doivent être utilisées pour éclairer les décisions pédagogiques, lesquelles doivent être meilleures que celles qui auraient été prises en l'absence de ces données. La longueur du cycle d'évaluation formative doit aussi être ajustée à la capacité du système à répondre aux données produites. Il ne sert à rien de recueillir des informations quotidiennement si les décisions qu'elles sont supposées éclairer sont prises une fois par mois (Wiliam et Thompson, 2007).

Les exemples qui répondent à cette définition ne relèvent pas tous de l'évaluation formative au sens de certaines autres définitions analysées plus haut. Ainsi, Cowie et Bell (1999), Looney (2005), Shepard (2007) et Kahl (2005) auraient sans doute hésité à employer le terme « formative » pour une évaluation qui serait éloignée du recueil des données. Le tour d'horizon de la littérature scientifique présenté ci-dessus confirme sans conteste que plus l'évaluation formative intervient tôt, plus est forte la probabilité de gains d'apprentissage et de gains importants. Toutefois, comme je l'ai mentionné dans une autre publication (Wiliam, 2009), il semble étrange de réserver le terme de « formative » aux évaluations qui ont un effet très important sur les résultats des élèves. Il serait plus pertinent à mon avis de qualifier l'évaluation de « formative » quand elle **forme** l'orientation des apprentissages futurs, tout en reconnaissant que cela implique des longueurs de cycle différentes, comme illustré au tableau 6.2.

Tableau 6.2. **Longueur des cycles dans l'évaluation formative**

| Type | Périodes visées | Longueur |
|-------------|---|---|
| Cycle long | Couvrant plusieurs périodes de notation ; trimestres, semestres, années | 4 semaines à 1 an |
| Cycle moyen | Pendant et entre les séquences d'enseignement | 1 à 4 semaines |
| Cycle court | Pendant et entre les cours | Au jour le jour : de 24 à 48 heures Minute par minute : de 5 secondes à 2 heures |

Source : Wiliam et Thompson (2007).

Évaluation formative : les principaux processus pédagogiques

Pour bien comprendre quelles sont les formes d'évaluation formative susceptibles d'être les plus efficaces, il convient de dépasser la définition fonctionnelle qui en est donnée et d'examiner plus en détails les processus sous-jacents. La métaphore des « systèmes » adoptée par Ramaprasad (1983) qui a servi de base à la définition de l'évaluation pour l'apprentissage adoptée par l'*Assessment Reform Group* (Groupe pour la réforme de l'évaluation, Broadfoot *et al.*, 2002), met en lumière trois processus pédagogiques clés visant à établir :

1. où en sont les élèves dans leurs apprentissages ;
2. où vont-ils ;
3. que doit-on faire pour les y amener.

Si de nombreuses stratégies d'évaluation formative mettent l'accent sur le rôle de l'enseignant, la définition adoptée ici reconnaît le rôle que les élèves et leurs pairs ont aussi à jouer. Le croisement de la dimension du processus (où en sont les élèves dans leurs apprentissages, où vont-ils et comment y parvenir) avec l'agent du processus (enseignant, pair, apprenant) permet d'établir une matrice de neuf cellules. Toutefois, si certaines cellules de cette matrice ont un sens en elles-mêmes, il peut s'avérer judicieux d'y associer l'examen d'autres cellules. Ainsi, si on considère le rôle des élèves dans la détermination du point où ils se trouvent dans leurs apprentissages et comment atteindre le but souhaité, on peut le représenter comme un processus visant à « inciter les élèves à s'approprier leur apprentissage », ce qui englobe plusieurs volets importants de l'apprentissage,

Tableau 6.3. **Stratégies d'évaluation formative en situation de classe**

| | <i>Où va l'apprenant ?</i> | <i>Où en est-il à cet instant précis ?</i> | <i>Comment y parvenir ?</i> |
|-------------------|---|---|---|
| <i>Enseignant</i> | Clarifier les intentions d'apprentissage, communiquer et fixer les critères de réussite (1) | Organiser de véritables discussions, des activités et des tâches qui produisent des données d'apprentissage (2) | Donner un feedback qui fait progresser les élèves (3) |
| <i>Pair</i> | Comprendre et partager les intentions d'apprentissage et les critères de réussite (1) | Inciter les élèves à être des personnes ressources l'un pour l'autre (4) | |
| <i>Apprenant</i> | Comprendre les intentions d'apprentissage et les critères de réussite (1) | Inciter les élèves à s'approprier leur apprentissage (5) | |

Source : Leahy, Lyon, Thompson et Wiliam, 2005.

notamment la métacognition (voir Schneider et Stern, ce volume). De même, le rôle des pairs dans la détermination du point où ils se trouvent et des moyens d'atteindre l'objectif fixé peut être présenté comme « inciter les élèves à être des personnes ressources l'un pour l'autre » (voir Barron et Darling Hammond, ce volume). Enfin, les trois cellules qui correspondent à la question « où vont-ils » peuvent être présentées comme « clarifier les intentions d'apprentissage, communiquer et fixer les critères de réussite ». Les neuf cellules peuvent ainsi se réduire aux cinq « stratégies de classe » de l'évaluation formative numérotées de 1 à 5 au tableau 6.3. Des précisions sur la base scientifique de ces cinq stratégies sont données dans Wiliam (2007) ; des informations plus précises sur le déploiement de ces stratégies par les enseignants au niveau de leur propre classe sont présentées dans Leahy, Lyon, Thompson et Wiliam (2005).

Évaluation formative et régulation des processus d'apprentissage

Dans la suite de ce chapitre, j'analyse comment la démarche d'évaluation formative exposée ici peut s'intégrer dans une perspective plus large de design pédagogique en s'intéressant plus particulièrement à la régulation des processus d'apprentissage (Perrenoud, 1991 ; 1998).

Ce cadre théorique donne des outils pour évaluer les actions de l'enseignant, des élèves et la situation d'apprentissage en termes de qualité de progression par rapport au but fixé. Comme le montrent Schneider et Stern (ce volume), l'enseignant ne crée pas l'apprentissage ; seuls les élèves peuvent le faire. De nombreux experts de l'éducation plaident pour que l'enseignant, qui joue traditionnellement le rôle du « sage sur l'éstrade » (« *sage on the stage* »), se transforme en « guide-accompagnateur » (« *guide on the side* »). Le danger de cette caractérisation est qu'elle est souvent interprétée comme une exonération des responsabilités de l'enseignant quant aux acquisitions effectives des élèves. Ce que je propose ici est de considérer que le rôle de l'enseignant est « d'organiser » un environnement d'apprentissage, tant du point de vue de sa conception que de son exploitation.

Un environnement d'apprentissage efficace est un environnement bien régulé, qui suscite l'implication des élèves. Les recherches toujours plus nombreuses dans le champ du développement cognitif montrent que le niveau d'implication dans des environnements faisant appel aux capacités cognitives a une incidence sur les résultats, mais aussi sur le quotient intellectuel (Dickens et Flynn, 2001 ; Mercer, Dawes, Wegerif et Sams, 2004). En outre, les environnements d'apprentissage efficaces doivent être, dans la mesure du possible, conçus de manière à favoriser ou étayer l'apprentissage visé (« régulation proactive »). Lorsque l'apprentissage visé n'a pas lieu, il est clair que des modifications doivent intervenir (« régulation interactive »). Enfin, l'enseignant peut aussi adopter une stratégie de « régulation rétroactive », par exemple lorsqu'il s'aperçoit qu'une séquence d'apprentissage pourrait être améliorée pour un groupe d'élèves à la lumière de son expérience avec les autres groupes.

La régulation proactive intervient en « amont » de la leçon proprement dite (c'est-à-dire avant que le cours ne commence). La régulation peut s'effectuer sans médiation, par exemple quand l'enseignant « n'intervient pas en personne, mais met en place une 'culture de la métacognition', des formes d'enseignement mutuel et des dispositifs de régulation des processus d'apprentissage pris en charge par des technologies ou incorporés aux dispositifs et aux situations didactiques » (Perrenoud, 1998, p. 100). Ainsi, la décision prise par l'enseignant de faire appel à des contextes réalistes en mathématiques peut offrir une source de régulation puisque les élèves sont à même d'évaluer dans quelle mesure leurs réponses sont raisonnables. Lorsqu'un enseignant développe chez les élèves des compétences de consultation et de soutien productif aux autres, on peut là encore parler de régulation proactive.

Dans d'autres cas, particulièrement s'il est difficile de prévoir comment les élèves vont réagir aux activités pédagogiques proposées, une régulation interactive des processus d'apprentissage peut être plus adaptée – par exemple en formulant des questions ou des activités qui suscitent des réactions de la part des élèves et sur lesquelles l'enseignant peut s'appuyer pour évaluer la progression des élèves et éventuellement procéder à des ajustements. Ces questions seront le plus souvent ouvertes et devront favoriser une réflexion poussée; elles sont en effet essentielles si l'on veut créer des situations d'apprentissages qui stimulent les élèves. Mais les questions fermées ont aussi un rôle à jouer. Des questions telles que « Le calcul est-il exact ou approché? », « Quel est le pH de 10 moles de NaOH? », ou encore « Votre masse serait-elle la même sur la lune? » sont bien des questions fermées avec une seule réponse possible. Elles ne sont toutefois pas dénuées d'intérêt car elles révèlent souvent chez les élèves des conceptions différentes de celles visées par l'enseignant (nombre d'élèves pensent que le calcul est approché, qu'un pH ne peut pas être supérieur à 14 et que la masse, comme le poids, dépend de la gravité).

La programmation « en amont » des bonnes questions à poser (voir ci-dessus) crée donc les conditions nécessaires pour permettre d'éventuellement modifier le déroulement des activités d'apprentissage « en aval » à la lumière des réponses des élèves. Ces « moments de contingence », c'est-à-dire les moments de la séquence pédagogique durant lesquels le cours peut partir dans des directions différentes en fonction des réponses des élèves, sont au cœur de la régulation des processus d'apprentissage. Et en effet, Black et Wiliam (2009, p. 6) déclarent que l'évaluation formative touche, par essence, à « la création et la mise à profit de 'moments de contingence' dans l'enseignement visant à réguler les processus d'apprentissage ». La théorie de l'évaluation formative est donc beaucoup plus étroite que la théorie globale de l'enseignement et de l'apprentissage, même s'il existe des liens forts entre les deux. Les stratégies mises en œuvre par les enseignants, les élèves et leurs pairs pour créer de tels moments de contingence et en tirer profit supposent en effet de prendre en considération le design pédagogique, le curriculum, la pédagogie, la psychologie et l'épistémologie.

Résumé

Ce chapitre a retracé les grands courants d'évolution du concept d'évaluation formative, mais le panorama ainsi dressé demeure inéluctablement très sélectif. Les premiers emplois de ce terme mettent surtout l'accent sur l'idée de feedback et les métaphores propres à la navigation, le feedback étant envisagé comme une mesure corrective visant à remettre l'apprenant sur le bon chemin. Depuis une centaine d'années, des milliers d'études, dans le sens littéral du terme, ont cherché à déterminer quelles étaient les formes de feedback susceptibles d'améliorer les processus d'apprentissage, et dans quelles proportions. Pourtant, ces études ne présentent qu'un intérêt limité, faute d'analyses théoriques solides sur les actions mêmes de feedback, sur les différents types d'apprentissage étudiés, et faute d'examen des effets à long terme. Les vingt dernières années ont vu émerger un intérêt considérable pour une évaluation formative qui ne serait plus isolée, mais qui serait une véritable composante des pratiques éducatives d'excellence en situation de classe. À ces fins, plusieurs définitions ont été avancées.

Ce chapitre présente une définition de l'évaluation formative qui met en relief le rôle de l'évaluation dans l'amélioration qualitative des choix pédagogiques et englobe toutes les définitions précédentes. Il dégage les implications de cette définition et propose de repenser l'évaluation formative comme impliquant cinq stratégies déterminantes :

1. Clarifier, communiquer et comprendre les intentions d'apprentissage et les critères de la réussite.
2. Concevoir des activités qui produisent des données sur les acquisitions.
3. Donner un feedback qui aide les apprenants à progresser.
4. Inciter les élèves à devenir des personnes ressources l'un pour l'autre.
5. Inciter les élèves à s'approprier leur apprentissage.

Enfin, il est suggéré que l'évaluation formative touche à la création et à la mise à profit de « moments de contingence » visant à réguler les processus d'apprentissage, ce qui permet de tracer une ligne de démarcation claire entre l'évaluation formative et les autres aspects du design pédagogique et de la pédagogie.

Notes

1. Dans un premier temps, ils ont sélectionné environ 3000 études scientifiques susceptibles d'intérêt. Ils ont ensuite éliminé toutes celles portant sur moins de 10 participants, celles qui ne comportaient pas de groupe témoin et celles dont le nombre de variables était insuffisant pour pouvoir calculer les effets. Restaient 131 publications, recensant 607 effets et couvrant 23 663 observations de 12 652 participants.
2. À partir d'une présélection dans des bases de données en ligne, 180 études significatives ont été retenues, 141 publications au total répondaient aux critères d'inclusion (103 articles publiés dans des revues, 24 ouvrages et chapitres de livres, 10 actes de conférence et 4 rapports de recherche).
3. En français, le terme « régulation » a un sens beaucoup plus étroit que le mot anglais « *regulation* ». Le terme « régulation », de plus en plus fréquent dans le discours anglophone sur les pratiques de classe, est employé dans l'acception française.

Bibliographie

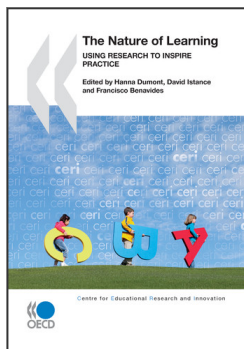
- Allal, L. et L.M. Lopez (2005), « L'évaluation formative de l'apprentissage : revue de publications en langue française », J. Looney (éd.), *L'évaluation formative : pour un meilleur apprentissage dans les classes secondaires*, Éditions OCDE, Paris, pp. 265-290.
- Ausubel, D.P. (1968), *Educational Psychology : A Cognitive View*, Holt, Rinehart and Winston, New York.
- Bangert-Drowns, R., C. Kulik, J. Kulik et M. Morgan (1991), « The Instructional Effect of Feedback in Test-Like Events », *Review of Educational Research*, vol. 61, n° 2, pp. 213-238.
- Black, P., C. Harrison, C. Lee, B. Marshall et D. Wiliam (2004), « Working Inside the Black Box : Assessment for Learning in the Classroom », *Phi Delta Kappan*, vol. 86, n° 1, pp. 9-21.
- Black, P. et D. Wiliam (1998a), « Assessment and Classroom Learning », *Assessment in Education : Principles Policy and Practice*, vol. 5, n° 1, pp. 7-73.
- Black, P. et D. Wiliam (2005), « Developing a Theory of Formative Assessment », J. Gardner (éd.), *Assessment and learning*, Sage, Londres, RU, pp. 81-100.
- Black, P. et D. Wiliam (2009), « Developing the Theory of Formative Assessment », *Educational Assessment, Evaluation and Accountability*, vol. 21, n° 1, pp. 5-31.
- Bloom, B.S. (1969), « Some Theoretical Issues Relating to Educational Evaluation », R.W. Tyler (éd.), *Educational Evaluation : New Roles, New Means : The 68th Yearbook of the National Society for the Study of Education (Part II)* (vol. 68, n 2, pp. 26-50), University of Chicago Press, Chicago, IL.
- Broadfoot, P., R. Daugherty, J. Gardner, W. Harlen, M. James et G. Stobart (1999), *Assessment for Learning : Beyond the Black Box*, School of Education, Université de Cambridge, Cambridge, RU.

- Broadfoot, P., R. Daugherty, J. Gardner, W. Harlen, M. James et G. Stobart (2002), *Assessment for Learning : 10 principes*, School of Education, Université de Cambridge, Cambridge, RU.
- Brookhart, S.M. (2004), « Classroom Assessment : Tensions and Intersections in Theory and Practice », *Teachers College Record*, vol. 106, n° 3, pp. 429-458.
- Brookhart, S.M. (2007), « Expanding Views about Formative Classroom Assessment : A Review of the Literature », J.H. McMillan (éd.), *Formative Classroom Assessment : Theory into Practice*, Teachers College Press, pp. 43-62.
- Cowie, B. et B. Bell (1999), « A Model of Formative Assessment in Science Education », *Assessment in Education : Principles Policy and Practice*, vol. 6, n° 1, pp. 32-42.
- Crooks, T.J. (1988), « The Impact of Classroom Evaluation Practices on Students », *Review of Educational Research*, vol. 58, n° 4, pp. 438-481.
- Deci, E.L. et R.M. Ryan (1994), « Promoting Self-Determined Education », *Scandinavian Journal of Educational Research*, vol. 38, n° 1, pp. 3-14.
- Dempster, F.N. (1991), « Synthesis of Research on Reviews and Tests », *Educational Leadership*, vol. 48, n° 7, pp. 71-76.
- Dempster, F.N. (1992), « Using Tests to Promote Learning : A Neglected Classroom Resource », *Journal of Research and Development in Education*, vol. 25, n° 4, pp. 213-217.
- Denvir, B. et M.L. Brown (1986a), « Understanding of Number Concepts in Low-Attaining 7-9 year olds : Part 1. Development of Descriptive Framework and Diagnostic Instrument », *Educational Studies in Mathematics*, vol. 17, n°1, pp. 15-36.
- Denvir, B. et M.L. Brown (1986b), « Understanding of Number Concepts in Low-Attaining 7-9 year olds : Part II, The Teaching Studies », *Educational Studies in Mathematics*, vol. 17, n° 2, pp. 143-164.
- Dickens, W. et J.R. Flynn (2001), « Heritability Estimates vs. Large Environmental Effects : The IQ Paradox Resolved », *Psychological Review*, vol. 108, n° 2, pp. 346-369.
- Elshout-Mohr, M. (1994), Feedback in Self-Instruction, *European Education*, vol. 26, n° 2, pp. 58-73.
- Forster, M. et G Masters (2004), « Bridging the Conceptual Gap between Classroom Assessment and System Accountability », M. Wilson (éd.), *Towards Coherence between Classroom Assessment and Accountability*, University of Chicago Press, Chicago.

- Fuchs, L.S. et D. Fuchs (1986), « Effects of Systematic Formative Evaluation – A Meta-Analysis », *Exceptional children*, vol. 53, n° 3, pp. 199-208.
- Harlow, L.L., S.A. Mulaik et J.H. Steiger (éd.) (1997), *What If There Were No Significance Tests ?* Lawrence Erlbaum Associates, Mahwah, NJ :
- Hattie, J. et H. Timperley (2007), « The Power of feedback », *Review of Educational Research*, vol. 77, n° 1, pp. 81-112.
- James, M. (1992), « Assessment for Learning », conférence annuelle de l'Association for Supervision and Curriculum Development (Session intitulée « Critique of Reforms in Assessment and Testing in Britain ») organisée à la Nouvelle Orléans, LA, University of Cambridge Institute of Education, Cambridge, RU.
- Kahl, S. (2005), « Where in the World are Formative Tests ? Right under Your Nose ! », *Education Week*, vol. 25, n° 4, p. 11.
- Kluger, A.N. et A. DeNisi (1996), « The Effects of Feedback Interventions on Performance : A Historical Review, A Meta-Analysis, and A Preliminary Feedback Intervention Theory », *Psychological Bulletin*, vol. 119, n° 2, pp. 254-284.
- Köller, O. (2005), « Formative Assessment in Classrooms : A Review of the Empirical German Literature », J. Looney (éd.), *Formative Assessment : Improving Learning in Secondary Classrooms*, Éditions OCDE, Paris, pp. 265-279.
- Leahy, S., C. Lyon, M. Thompson et D. Wiliam (2005), « Classroom Assessment : Minute-by-Minute and Day-by-Day », *Educational Leadership*, vol. 63, n° 3, pp. 18-24.
- Looney, J. (éd.) (2005), *Formative Assessment : Improving Learning in Secondary Classrooms*, Éditions OCDE, Paris.
- Meisels, S.J., S. Atkins-Burnett, Y. Xue, J. Nicholson, D.D. Bickel et S.-H. Son (2003), « Creating a System of Accountability : The impact of Instructional Assessment on Elementary Children's Achievement Test Scores », *Education Policy Analysis Archives*, vol. 11, n° 9.
- Mercer, N., L. Dawes, R. Wegerif et C. Sams (2004), « Reasoning as a Scientist : Ways of Helping Children to Use Language to Learn Science », *British Educational Research Journal*, vol. 30, n° 3, pp. 359-377.
- Natriello, G. (1987), « The Impact of Evaluation Processes on Students », *Educational Psychologist*, vol. 22, n° 2, pp. 155-175.

- Nyquist, J.B. (2003), *The Benefits of Reconstructing Feedback as a Larger System of Formative Assessment : A Meta-Analysis*, Mémoire de master en sciences non publié, Université Vanderbilt.
- Perrenoud, P. (1991), « Towards a Pragmatic Approach to Formative Evaluation », P. Weston (éd.), *Assessment of Pupils' Achievement : Motivation and School Success*, Amsterdam : Swets & Zeitlinger, pp. 77-101.
- Perrenoud, P. (1998), « From Formative Evaluation to a Controlled Regulation of Learning Towards a Wider Conceptual Field », *Assessment in Education : Principles Policy and Practice*, vol. 5, n° 1, pp. 85-102.
- Ramaprasad, A. (1983), « On the Definition of Feedback », *Behavioural Science*, vol. 28, n° 1, pp. 4-13.
- Sadler, D.R. (1989), « Formative Assessment and the Design of Instructional Systems », *Instructional Science*, vol. 18, n° 2, pp. 119-44.
- Shepard, L.A. (2007), « Formative Assessment : Caveat Emptor », C.A. Dwyer (éd.), *The Future of Assessment : Shaping Teaching and Learning*, Lawrence Erlbaum Associates, Mahwah, NJ, pp. 279-303.
- Shepard, L., K. Hammerness, L. Darling-Hammond et F. Rust (2005), « Assessment », L. Darling-Hammond et J. Bransford (éd.), *Preparing Teachers for a Changing World : What Teachers Should Learn and Be Able to Do*, CA : Jossey-Bass, San Francisco, CA, pp. 275-326.
- Shute, V.J. (2008), « Focus on Formative Feedback », *Review of Educational Research*, vol. 78, n° 1, pp. 153-189.
- Torrance, H. et J. Pryor (1998), *Investigating Formative Assessment*, Open University Press, Buckingham, RU.
- Tunstall, P. et C.V. Gipps (1996a), « Teacher Feedback to Young Children in Formative Assessment : A Typology », *British Educational Research Journal*, vol. 22, n° 4, pp. 389-404.
- Tunstall, P. et C.V. Gipps (1996b), « How Does Your Teacher Help You to Make Your Work Better? Children's Understanding of Formative Assessment », *The Curriculum Journal*, vol. 7, n°2, pp. 185-203.
- Vinner, S. (1997), « From Intuition to Inhibition – Mathematics, Education and Other Endangered Species », E. Pehkonen (éd.), *Actes de la 21^e Conférence du groupe international Psychology of Mathematics Education*, Université d'Helsinki, Centre de formation et de recherches de Lahti, Finlande, vol. 1, pp. 63-78.
- Wiener, N. (1948), *Cybernetics, or the Control and Communication in the Animal and the Machine*, John Wiley, New York, NY.

- Wiliam, D. (2007), « Keeping Learning on Track : Classroom Assessment and the Regulation of Learning », F.K. Lester Jr (éd.), *Second Handbook of Mathematics Teaching and Learning*, Information Age Publishing, Greenwich, CT, pp. 1053-1098.
- Wiliam, D. (2009), « An Integrative Summary of the Research Literature and Implications for a New Theory of Formative Assessment », H.L. Andrade et G.J. Cizek (éd.), *Handbook of Formative Assessment*, Routledge, Taylor and Francis, New York.
- Wiliam, D. et P.J. Black (1996), « Meanings and Consequences : A Basis for Distinguishing Formative and Summative Functions of Assessment? », *British Educational Research Journal*, vol. 22, n° 5, pp. 537-548.
- Wiliam, D. et M. Thompson (2007), « Integrating Assessment with Instruction : What Will it Take to Make it Work? », C.A. Dwyer (éd.), *The Future of Assessment : Shaping Teaching and Learning*, Lawrence Erlbaum Associates, Mahwah, NJ, pp. 53-82.



Extrait de :
The Nature of Learning
Using Research to Inspire Practice

Accéder à cette publication :
<https://doi.org/10.1787/9789264086487-en>

Merci de citer ce chapitre comme suit :

William, Dylan (2010), « Le rôle de l'évaluation formative dans les environnements d'apprentissage efficaces », dans Hanna Dumont, David Istance et Francisco Benavides (dir. pub.), *The Nature of Learning : Using Research to Inspire Practice*, Éditions OCDE, Paris.

DOI: <https://doi.org/10.1787/9789264086944-8-fr>

Cet ouvrage est publié sous la responsabilité du Secrétaire général de l'OCDE. Les opinions et les arguments exprimés ici ne reflètent pas nécessairement les vues officielles des pays membres de l'OCDE.

Ce document et toute carte qu'il peut comprendre sont sans préjudice du statut de tout territoire, de la souveraineté s'exerçant sur ce dernier, du tracé des frontières et limites internationales, et du nom de tout territoire, ville ou région.

Vous êtes autorisés à copier, télécharger ou imprimer du contenu OCDE pour votre utilisation personnelle. Vous pouvez inclure des extraits des publications, des bases de données et produits multimédia de l'OCDE dans vos documents, présentations, blogs, sites Internet et matériel d'enseignement, sous réserve de faire mention de la source OCDE et du copyright. Les demandes pour usage public ou commercial ou de traduction devront être adressées à rights@oecd.org. Les demandes d'autorisation de photocopier une partie de ce contenu à des fins publiques ou commerciales peuvent être obtenues auprès du Copyright Clearance Center (CCC) info@copyright.com ou du Centre français d'exploitation du droit de copie (CFC) contact@cfcopies.com.