« L'évaluation pour les apprentissages » : d'une expérimentation locale à une politique nationale

"Assessing for learning" from local experiments to national policy

*"La evaluación para los aprendizajes": de una experimentación local a una
política nacional*

*„Die Beurteilung für das Erlernen“: vom Lokalexperiment bis hin zur
Nationalpolitik*

Gordon Stobart

Édition électronique

URL : <http://journals.openedition.org/rfp/2657>

DOI : 10.4000/rfp.2657

ISSN : 2105-2913

Éditeur

ENS Éditions

Édition imprimée

Date de publication : 15 mars 2011

Pagination : 41-48

ISBN : 978-2-7342-1203-4

ISSN : 0556-7807

Référence électronique

Gordon Stobart, « « L'évaluation pour les apprentissages » : d'une expérimentation locale à une politique nationale », *Revue française de pédagogie* [En ligne], 174 | janvier-mars 2011, mis en ligne le 15 mars 2015, consulté le 30 avril 2019. URL : <http://journals.openedition.org/rfp/2657> ; DOI : 10.4000/rfp.2657

« L'évaluation pour les apprentissages » : d'une expérimentation locale à une politique nationale

Gordon Stobart

L'évaluation pour les apprentissages est une approche bien développée aujourd'hui au Royaume-Uni et elle fait partie des instructions officielles. Elle a montré sa pertinence et fait l'objet d'une reconnaissance internationale. Mais la façon dont cette expérimentation est mise en œuvre et interprétée par les responsables de l'éducation anglais n'est pas sans poser des problèmes : mauvaise compréhension des principes fondateurs et distorsion des pratiques. L'obligation de résultats de recherche dans une politique cohérente et intégrée peut se heurter à un certain nombre de défis lorsqu'elle est conduite de façon trop mécanique et superficielle (en ignorant notamment les réalités du contexte des établissements) et lorsqu'elle sous-estime l'effort d'engagement des enseignants et des élèves dans les processus d'évaluation et d'apprentissage.

Descripteurs (TESE) : évaluation formative, processus d'apprentissage, niveau national, évaluation du système éducatif, transfert des connaissances, politique en matière d'éducation.

INTRODUCTION

Cet article traite de l'appropriation et de la capitalisation des expérimentations pédagogiques locales par les agences centrales et les responsables gouvernementaux, ainsi que de leur impact sur le système éducatif. Dans certaines circonstances, un tel processus subit des ajustements et des modifications notables qui sont la conséquence d'une relative complexité de l'agenda politique, au détriment des bénéfices potentiels, mais aussi des crédits supplémentaires qu'exige ce type de transfert et de généralisation. En Angleterre, le gouvernement a généralisé une expérimentation

populaire dans le domaine de l'évaluation : l'évaluation pour les apprentissages (*assessment for learning*). Elle s'inscrit dans une forte culture de l'obligation de résultats qui vise à améliorer les résultats des élèves aux évaluations nationales et à atteindre les objectifs fixés au regard de standards nationaux, ce qui explique certains ajustements opérés par rapport aux intentions originelles de ceux qui avaient conçu ce dispositif expérimental. L'article présente une discussion des avantages et des inconvénients de ces expérimentations de terrain lorsque les responsables politiques et les financeurs se les approprient à d'autres fins.

L'ACTION PUBLIQUE ET LA CULTURE D'OBLIGATION DE RÉSULTATS

En Angleterre, les services publics développent une culture d'obligation de résultats qui repose sur des objectifs fixés par le gouvernement afin d'élever la qualité des services. Parmi ces objectifs, les hôpitaux doivent par exemple réduire les listes d'attente pour les opérations chirurgicales et le délai permettant de voir un docteur en cas d'urgence. De même, les compagnies de trains et de bus font face à une obligation de ponctualité et sont redevables d'une amende en cas de non-respect. Dans le domaine de l'éducation, les objectifs portent le plus souvent sur la réussite aux tests et aux examens nationaux. Les établissements scolaires, les collectivités locales et le gouvernement lui-même sont tenus de respecter des objectifs et d'afficher la progression de leurs résultats d'une année sur l'autre. C'est ainsi que le ministère a établi des objectifs de performance dans les évaluations nationales pour les élèves de 11 ans, à la fin de l'enseignement primaire. Le seuil fixé (c'est-à-dire 85 % d'élèves atteignant le niveau 4 en anglais et en mathématiques, niveau attendu d'un élève de 11 ans) n'ayant pas été atteint, cela a même entraîné la démission d'un ministre de l'Éducation anglais. Des objectifs sont également déterminés pour les établissements de l'enseignement secondaire, à partir des examens nationaux destinés aux élèves de 16 ans. Les établissements dont les résultats sont en deçà des attentes sont susceptibles d'être inspectés, voire fermés dans le cadre de la politique qualifiée de « défi national », quand bien même leurs effectifs seraient constitués en majorité d'élèves défavorisés et en échec scolaire.

L'un des effets secondaires introduits par la culture du résultat est que les services publics cherchent désormais à contourner le système par tous les moyens, afin de remplir leurs objectifs. Les hôpitaux ont ainsi mis en œuvre un grand nombre d'opérations bénignes et non urgentes afin de réduire leur liste d'attente, retardant d'autant les actes plus lourds de chirurgie. Les compagnies ferroviaires ont modifié leurs horaires en rallongeant les temps de trajet afin d'éviter les pénalités liées aux retards. Les établissements scolaires sont passés maîtres dans l'art de choisir les enseignements susceptibles de gonfler leurs résultats aux examens (par exemple les enseignements professionnels dans les technologies de l'information et de la communication). Le gouvernement a même renforcé cette tendance par l'octroi de crédits supplémentaires destinés aux élèves proches

de la réussite, dans l'espoir que cette attention et les ressources supplémentaires amélioreraient leurs résultats et les aideraient ainsi à atteindre les objectifs. Paradoxalement les élèves qui n'avaient guère de chances de réussir ne bénéficiaient pas de ces avantages, forme perverse s'il en est de « tri éducatif » (Gillborn & Youdell, 2000). O'Neill (2002) et Stobart (2008) en ont livré un compte rendu détaillé en étudiant l'impact négatif de ces objectifs stratégiques.

Malgré tout, l'un des aspects positifs de cette volonté politique de relever le niveau d'exigence est le soutien aux expérimentations pédagogiques susceptibles d'améliorer les résultats des élèves aux tests. En Angleterre, cela s'est traduit parfois par un enseignement prescriptif des mathématiques et de l'anglais en termes de contenus, de méthodes pédagogiques et de séquences d'apprentissage (ironie du sort, l'Angleterre semble reproduire le stéréotype français selon lequel chaque élève étudierait la même chose au même moment). La recherche de l'amélioration des compétences des élèves peut s'appuyer sur des expérimentations locales réussies et leur extension à l'échelon national. Citons par exemple l'apprentissage précoce de la lecture, qui a fait l'objet d'après débats au Parlement quant à la meilleure méthode à adopter pour sa mise en œuvre. Ces débats faisaient pourtant suite au bilan convaincant établi par une collectivité locale en Écosse, qui s'était engagée dans cette expérimentation. Les développements de l'intelligence émotionnelle et les activités de pratique réflexive parmi les élèves font aussi partie de ces expérimentations locales.

Cet article décrit à la fois un projet expérimenté sur le terrain, visant à améliorer l'évaluation des élèves en classe, et l'analyse de ses effets secondaires après que le dispositif a été repris par les décideurs politiques. Cette étude de cas porte sur l'évaluation pour les apprentissages en Angleterre, une méthode d'évaluation formative qui s'est développée au cours des dix dernières années, depuis la célèbre publication du livret *Inside the black box* (Black & William, 1998) et de l'ouvrage *Assessment for learning: Beyond the black box* (Assessment Reform Group, 1999). Cette expérimentation a été reprise par les pouvoirs publics et elle a constitué le cœur des stratégies nationales mises en œuvre dans les enseignements primaire et secondaire britannique. En 2008, le ministère de l'Éducation a lancé le programme *Évaluation pour les apprentissages*, qui a été financé à hauteur de 150 millions de livres sterling (Department of Children, Schools and Families, 2008). C'est ce programme qui est ici présenté et analysé.

UNE DÉFINITION DE L'ÉVALUATION FORMATIVE

La littérature anglo-saxonne (entre autres) établit une différence entre évaluation sommative et évaluation formative. La première constitue une photographie des acquis des élèves à un moment donné, par exemple suite à un test de fin d'année. La seconde entend contribuer au processus continu d'apprentissage : elle « met en lumière » les apprentissages des élèves. Cependant, alors que cette dernière notion fut inventée aux États-Unis dans les années soixante, son sens a évolué et désigne désormais des méthodes très variées. Dans leur synthèse de la littérature française sur l'évaluation formative, Allal et Mottier Lopez (2005) opèrent une distinction entre trois types de réponse formative aux informations provenant des évaluations (elles parlent de régulation) :

- la régulation interactive part des interactions entre l'élève et les autres personnes et éléments matériels présents dans la salle de classe. Elle est largement ancrée dans les pratiques quotidiennes et conduit à une adaptation continue de l'apprentissage, notamment par le suivi individualisé des élèves et le *feedback*. Elle coïncide avec des approches socioculturelles et constructivistes de l'apprentissage ;
- la régulation rétroactive concerne l'évaluation formative conduite après une séquence de cours, souvent par l'intermédiaire d'un test. Elle consiste à traiter les difficultés d'apprentissage qui ont été repérées parmi les élèves. Ce modèle du « tester et remédier » demeure prédominant aux États-Unis et il est le reflet de méthodes comportementalistes déjà anciennes ;
- la régulation proactive intervient lorsque les résultats des élèves aux tests amènent des changements dans les méthodes d'enseignement. En France, il s'agit de diversifier les activités pour répondre aux besoins des élèves. Les enseignants peuvent être amenés à modifier leurs pratiques, mais avec un décalage parce que les résultats des tests arrivent trop tard. Ils ajustent alors les contenus et les pratiques d'enseignement au groupe d'élèves suivant.

Si ces trois types de régulation constituent bien le répertoire principal d'un enseignant, l'équilibre entre les trois fait appel à des différences d'interprétation. Au cœur de ces différences figure en bonne place la conception de l'apprentissage : est-il centré sur l'élève ou sur l'enseignant ? Dans les méthodes rétroactives et proactives, les enseignants sont les

principaux apprenants, dans la mesure où ils ajustent leurs pratiques. La méthode interactive est davantage focalisée sur l'élève, le rôle de l'enseignant consistant à remettre les clés de l'apprentissage à l'élève au fur et à mesure qu'il gagne en autonomie. De telles différences de points de vue sont parfois source de tension. Par exemple, en Angleterre, les responsables de l'éducation s'intéressent aux enseignants puisqu'ils sont la cible de leur politique ; on met l'accent sur la façon dont ils devraient modifier leur pratique d'enseignement (d'où le postulat selon lequel les enseignants doivent d'abord dominer le processus d'apprentissage). Cependant ce type de politique peut constituer un frein aux méthodes centrées sur l'apprenant, comme l'évaluation pour les apprentissages, qui cherche à impliquer un peu moins les enseignants et davantage les élèves. Quant aux approches plus commerciales, sans surprise, elles placent les batteries de tests au cœur de l'évaluation formative puisqu'elles cherchent à vendre des outils de diagnostic aux établissements scolaires sous une forme ou sous une autre.

L'ÉVALUATION POUR LES APPRENTISSAGES

L'évaluation pour les apprentissages est une méthode d'évaluation formative culturellement ancrée dans les pays anglo-saxons comme le Royaume-Uni, la Nouvelle-Zélande, l'Australie et le Canada. Il s'agit d'une interprétation culturelle spécifique de l'évaluation formative qui s'inscrit dans ce que Mottier Lopez et Allal appellent la régulation interactive. Le terme est préféré à celui d'évaluation formative car il est plus clair quant à son objectif : contribuer directement au processus d'apprentissage des élèves. Le terme « évaluation pour les apprentissages » a été introduit en raison des nombreuses méprises qu'entraîne le mot « formatif ». Parmi elles subsiste l'idée que les tests en classe, destinés avant tout à suivre la progression des élèves, font partie de l'évaluation formative. Étant donné leur objectif, le terme de « mini-évaluations sommatives » conviendrait mieux, dans la mesure où les informations recueillies ne servent pas à ajuster directement les pratiques pédagogiques des enseignants. Il en est de même de la notation du travail en classe, souvent décrite comme formative, alors qu'elle sert en réalité à asseoir des jugements sommatifs. Néanmoins la relation entre formatif et sommatif est loin d'être clairement définie dans la littérature de recherche. L'évaluation pour les apprentissages est un mode d'évaluation intégré

dans le processus d'apprentissage des élèves. En voici cinq points clés à la simplicité trompeuse (Assessment Reform Group, 1999, p. 4-5) :

- la participation active des élèves à leur apprentissage ;
- l'apport d'un *feedback* efficace pour les élèves ;
- l'adaptation de l'enseignement pour prendre en compte les résultats de l'évaluation des élèves ;
- la nécessité pour les élèves d'être en mesure de s'auto-évaluer ;
- la reconnaissance de l'influence profonde de l'évaluation sur la motivation et l'estime de soi des élèves, deux conditions essentielles à leur apprentissage.

Voici la définition qu'en propose l'Assessment Reform Group : « Processus qui consiste à chercher et interpréter les ressources utiles aux élèves et aux enseignants, à déterminer où en sont les élèves dans leur apprentissage, où ils doivent aller et le meilleur moyen d'y parvenir. » (Assessment Reform Group, 2002) Cette méthode correspond finalement à un enseignement de qualité surtout si, par « ressources », on entend le recours à une large palette d'informations plutôt qu'à des tests formels ou informels. L'accent mis sur l'évaluation résulte d'une volonté délibérée des membres du groupe de réaffirmer l'objectif essentiel – mais aujourd'hui menacé – de l'évaluation. À une époque dominée par l'évaluation sommative, cette méthode cherche à rééquilibrer l'usage de l'évaluation en l'*intégrant* au processus d'apprentissage, plutôt que d'en faire un objet externe.

L'évaluation pour les apprentissages correspond à des pratiques pédagogiques spécifiques qui doivent prendre en compte les *intentions d'apprentissage* des élèves et fixer *a priori* des *critères de réussite*, c'est-à-dire expliciter ce qui est enseigné et préciser la définition d'une performance réussie. Elle s'attache par exemple à la manière dont les enseignants formulent leurs questions aux élèves, à ce qu'ils leur laissent suffisamment de temps pour répondre et à favoriser les réponses collaboratives de leur part. Cette démarche exige des enseignants qu'ils affinent leurs questions pour mieux révéler « à quel stade les élèves en sont dans leur apprentissage » et ce qu'il en est de leurs erreurs. Récemment, le dialogue entre l'enseignant et ses élèves, bien plus encore que la formulation des questions, a été mis au centre du dispositif, comme avec l'usage des *feux de signalisation* : les élèves, souvent placés en groupe, signalent s'ils ont compris le cours (en vert), s'ils ont des doutes (en orange) ou s'ils n'ont pas compris (en rouge). Le

feedback est alors perçu comme le dispositif essentiel pour réduire l'écart entre le stade actuel d'apprentissage des élèves et le niveau à atteindre. La complexité de ce processus est de plus en plus reconnue et, bien souvent, on considère à tort que le dispositif utilisé est un *feedback*, alors qu'il ne favorise pas réellement l'acquisition des connaissances des élèves (Kluger & Denisi, 1996 ; Hattie & Timperley, 2007). Dans la pratique, l'accent doit être mis sur les commentaires fait par les élèves à partir des tâches qu'ils exécutent, bien plus que sur les notes qui sont, comme l'a écrit Sadler, « trop profondément codées » (Sadler, 1998) pour donner des informations précises sur ce qu'il convient de faire par la suite. Une des découvertes les plus stimulantes est que le *feedback* trop centré sur l'ego au détriment de la tâche (par exemple, quand on dit à l'élève : « Tu me déçois ») est nuisible à un apprentissage efficace. Cela pose la question du rôle des affects (comme les compliments) dans l'évaluation formative.

L'évaluation pour les apprentissages à l'échelle de l'établissement

Cette formule d'« évaluation pour les apprentissages » est le fruit de la collaboration entre un groupe informel d'universitaires et des établissements avec lesquels ils ont travaillé, souvent en lien étroit, pour développer cette méthode particulière dans l'environnement de la salle de classe. Cette approche de l'évaluation est désormais suffisamment établie au Royaume-Uni pour être traitée comme un mouvement à part entière. Elle est aujourd'hui reprise dans le monde entier. Si le terme existe depuis environ une quinzaine d'années, sa popularité récente est liée au livret *Inside the black box* de Black et William (1998), dont plus de 50 000 exemplaires ont été vendus à travers le monde. Ce livret part de leur synthèse systématique de la littérature de recherche sur l'évaluation en classe et porte le souci d'ouvrir cette boîte noire et d'analyser ce qui s'y passe (Black & William, 1998). Plusieurs livres ont été publiés depuis, dont l'ouvrage *Assessment for learning: Beyond the black box*, rédigé par l'Assessment Reform Group (Groupe de réforme de l'évaluation). Celui-ci a également bénéficié d'un large écho et conduit à la généralisation de la méthode à l'ensemble du système éducatif anglais. Son influence grandissante se mesure aussi par la reprise de ses principales idées clés dans les stratégies nationales en éducation, au Royaume-Uni comme dans d'autres pays.

Auto-évaluation et évaluation par les pairs

L'un des objectifs clés de l'évaluation pour les apprentissages est de favoriser une culture de classe où les élèves sont en mesure de juger à la fois la qualité de leur travail et de celui de leurs pairs, afin de comprendre ce qu'implique pour eux un apprentissage efficace. La logique de cette démarche est de faire en sorte que les élèves, pour évaluer leur propre travail, sachent appréhender une performance réussie et déterminer à quel stade ils en sont dans leur apprentissage. Ces compétences sont au fondement de l'autorégulation, considérée comme la condition pertinente d'un apprentissage efficace. L'enseignant change alors de rôle pour partager ses connaissances sur l'évaluation avec ses élèves. Cette adaptation nécessite la hiérarchisation des pratiques efficaces dans la classe (« Pourquoi ce travail-ci est-il meilleur que celui-là ? ») et le souci constant de fournir un *feedback* aux élèves.

L'attrait de cette méthode pour les enseignants

L'évaluation pour les apprentissages est populaire auprès des enseignants, du fait de son aspect pratique et de son ancrage dans la vie de la classe : ce sont quelques techniques qu'il est facile de tester en classe. Certaines ne perturbent guère le cours – la gestion du temps d'attente des réponses des élèves par les enseignants peut se pratiquer aisément dans l'intimité de la salle de classe –, bien qu'elles puissent conduire à des changements profonds. Les interactions et l'environnement de classe les plus propices à un apprentissage efficace sont bien révélés par cette méthode. En 2007, Crooks, un important théoricien de l'évaluation formative, recensait les principaux facteurs d'une utilisation efficace de l'évaluation : celle-ci passe par l'interaction de l'enseignant avec ses élèves et par le développement d'une culture de confiance dans la classe (Crooks, 2007).

L'appropriation de l'évaluation pour les apprentissages par les décideurs politiques

L'évaluation pour les apprentissages est un mouvement parti du terrain, sur la base d'une recherche-action (voir par exemple, Black & William, 2003) et de plusieurs séances de développement professionnel continu pour les enseignants (voir par exemple, Clarke, 2001). La méthode a été bien acceptée par les enseignants et les établissements scolaires, d'où l'intérêt manifesté par les collectivités et plusieurs gouvernements locaux, qui ont par la suite encouragé plusieurs expérimentations de même type. Ainsi,

en Écosse, certains établissements étaient accompagnés dans ces expérimentations par l'attribution de crédits supplémentaires, et par des regroupements locaux d'enseignants. Au pays de Galles, le gouvernement s'est montré plus directif en mettant en place des projets pilotes. L'Angleterre, quant à elle, a associé expérimentations locales et projets centralisés : l'évaluation pour les apprentissages a été incluse dans les stratégies nationales pour les enseignements primaire et secondaire. La mise en œuvre du dispositif a nécessité la rédaction de manuels et le développement de formations « clés en main », où le formateur était tenu de dispenser la formation selon un format prescrit à l'avance (présentation par items indiquant la marche à suivre). Le gouvernement finançait aussi les établissements volontaires pour qu'ils intègrent l'évaluation pour les apprentissages dans leurs programmes stratégiques. À ce stade, l'appropriation de la méthode par les décideurs politiques fut perçue par les gens de terrain comme une reconnaissance utile des revendications qu'ils portaient, malgré une présentation souvent convenue des pratiques pédagogiques en classe et l'omission de quelques éléments clés comme l'auto-évaluation, l'évaluation par les pairs ou l'idée qu'il fallait donner la préférence aux commentaires écrits des élèves par rapport aux notes de leurs copies. L'approche moins prescriptive adoptée par l'Écosse était considérée comme le modèle le plus approprié (Gardner, Harlen, Hayward *et al.*, 2010).

La stratégie d'évaluation pour les apprentissages annoncée par le gouvernement anglais dès 2008 constitue la parfaite illustration des difficultés rencontrées par les décideurs politiques pour adopter des expérimentations locales. En apparence, elle se présente comme une stratégie nationale bien financée encourageant les établissements à développer leurs pratiques d'évaluation formative, conformément à la définition qu'en a donnée l'Assessment Reform Group. Néanmoins d'autres enjeux politiques viennent interférer dans le dispositif, à la suite des changements du système national de tests sommatifs des compétences des élèves. Cela est dû au retrait par le gouvernement de certains tests et à la controverse qu'a suscitée le maintien des autres. En effet les tests nationaux destinés aux élèves de 14 ans ont été retirés soudainement, à la suite de la polémique qui a affecté les scores des élèves en 2008 : l'agence américaine de test (Educational Testing Service) choisie par le gouvernement n'est pas parvenue à publier les résultats des élèves dans les temps. En termes d'obligation de résultats, ça ne posait guère de problème dans la mesure où les objectifs clés du gouvernement sont

fondés sur les examens nationaux destinés aux élèves de 16 ans. En revanche, toutes les mesures des performances scolaires des élèves de 14 ans en anglais, mathématiques et sciences reposaient sur les seules évaluations des élèves réalisées par leurs enseignants. Quelles mesures fallait-il donc prendre pour les rendre plus précises et plus fiables ? La même question se posait pour les évaluations données au niveau du premier cycle de l'enseignement secondaire au moment où, à l'échelle nationale, de nouveaux tests pilotes se mettaient en place, pour lesquels les enseignants devaient désigner un seul élève par niveau de scolarité (alors que jusque-là tous les élèves effectuaient le même test sur plusieurs cycles de scolarité). Il était important que les enseignants soient capables d'évaluer précisément le niveau atteint par leurs élèves.

Le gouvernement s'est alors saisi de ces inquiétudes et a mis en place un programme intitulé *Évaluer les progrès des élèves (Assessing Pupil Progress)*, qui cherchait à standardiser les évaluations conçues par les enseignants en fournissant des cadres d'évaluation et des ressources types pour leur permettre de mieux calibrer les épreuves. Mais les enseignants ont estimé que ces nouvelles exigences allaient alourdir encore un peu plus leur charge de travail. Dès lors, comment pouvait-on leur faire accepter ce nouveau programme ? Il fallait utiliser l'évaluation pour les apprentissages. Cette méthode, populaire dans les établissements, devait aider à « faire passer la pilule ». En fait, je préfère encore la métaphore du cheval de Troie : ce qui ressemble d'abord à un cadeau du gouvernement est intégré progressivement dans les établissements avant de devenir beaucoup moins attrayant pour les enseignants. À cette fin, des changements subtils sont apparus. La partie intitulée « Savoir où en sont les élèves », correspondant à la définition originelle, se transforme en : « Une bonne évaluation passe par une évaluation précise : connaître les niveaux d'exigence, juger du travail des élèves de façon appropriée et concevoir des évaluations précises en rapport avec les niveaux fixés par les programmes nationaux. » (Department of Children, Schools and Families, 2008, p. 5) Dans la même veine, la partie « Objectifs à atteindre » revient à préciser quel niveau (il y en a huit) des programmes nationaux l'élève doit atteindre, au lieu de mesurer précisément ses connaissances et compétences acquises. Au-delà de son caractère particulier, cette illustration de l'appropriation et de la transformation par les décideurs politiques d'une expérimentation issue du terrain montre la façon dont ces derniers reprennent une expérimentation populaire pour l'utili-

ser à des fins autres qu'une simple généralisation. Étant donné les objectifs et les agendas multiples des gouvernements et de leurs agences, les personnes qui se lancent dans ce genre d'initiative devraient se méfier par avance des changements subtils qui peuvent en saper l'esprit originel.

LES AVANTAGES D'UN SOUTIEN POLITIQUE

Précédemment, j'ai présenté les risques potentiels liés à la reprise politique d'une expérimentation locale. Considérons maintenant ses bénéfices éventuels. Ils peuvent être regroupés sous quatre intitulés : la légitimité, les ressources supplémentaires, le soutien de l'équipe de direction et le développement d'une expertise locale.

Légitimité

Le gouvernement encourage officiellement l'essai de cette nouvelle méthode. L'assentiment des autorités publiques est déterminant dans un pays où la prise de risques peut être fortement pénalisée. En Angleterre, l'agence de l'inspection des établissements (l'OFSTED) joue un rôle central dans le contrôle de conformité aux instructions officielles et n'hésite pas à sanctionner les établissements qui s'en écartent. D'ailleurs, ces derniers se garderaient bien de prendre des initiatives sans l'approbation des inspecteurs de cette agence. Leur feu vert constitue dans ce cas une garantie et il permet de changer les pratiques pédagogiques dans la classe. Dans le cas de l'évaluation pour les apprentissages, la communauté universitaire a aussi accordé un blanc-seing au dispositif. La grande synthèse de la littérature de recherche opérée par Black et William sur l'évaluation des élèves en classe (Black & William, 1998) a constitué le point de départ de recherches sur les pratiques d'enseignement et les situations d'apprentissage. Cette synthèse scientifique fut ensuite reprise dans un ouvrage de format plus restreint à destination des enseignants et des décideurs politiques. Les enseignants et les établissements se sont alors lancés dans cette nouvelle méthode. En Écosse, les responsables de l'éducation ont utilisé ces travaux pour développer leur propre système d'évaluation pour les apprentissages. L'influence de cette recherche sur la politique a été explorée plus en détail dans les écrits de Daugherty (2007).

De façon générale, les enseignants doivent avoir de bonnes raisons pour changer leurs pratiques traditionnelles dans les établissements scolaires. À quoi bon tenter quelque chose de nouveau si c'est pour rencontrer plus de problèmes qu'avec les méthodes déjà connues, même s'il est avéré que celles-ci ne sont pas vraiment efficaces ? Des encouragements provenant des chercheurs ou des décideurs politiques peuvent inciter les enseignants les plus expérimentés à quelques changements de pratiques.

Des ressources supplémentaires

La reprise politique d'une expérimentation locale génère parfois une injection d'argent frais et des ressources supplémentaires. Ces expérimentations sont souvent en effet sous-financées et reposent sur la mise en commun de ressources aux origines diverses. De plus, la reprise par une agence mieux dotée sur le plan financier permet d'envisager une approche plus intégrée ou systémique du dispositif. Les missions confiées à l'Assessment Reform Group en constituent un bon exemple. Il s'agit d'un regroupement informel de huit à dix universitaires experts dans le domaine de l'évaluation. Ils n'ont ni bureau, ni budget dédié, ni soutien administratif direct. Le seul financement provient de bourses de recherche, notamment de la fondation Nuffield en Angleterre, de la vente de leurs ouvrages et des droits d'auteur. C'est à partir de la bourse octroyée par la fondation Nuffield qu'a été commandée une revue systématique de la recherche internationale encadrée par Black et William (1998). L'adoption de l'évaluation pour les apprentissages par diverses agences gouvernementales au Royaume-Uni a permis ensuite une diffusion bien plus large, soit sous forme numérique, soit sous forme papier. C'est ainsi que fut expédiée une brochure intitulée *Évaluation pour les apprentissages : les dix principes* à presque tous les établissements publics du Royaume-Uni. En parallèle, le gouvernement anglais a financé des études pilotes sur l'évaluation pour les apprentissages, le plus souvent sous la forme de recherche-action dans les établissements scolaires. En Angleterre, ce dispositif s'inscrivait dans le cadre plus large des stratégies nationales pour la littératie et la numératie. En Écosse, il a fait partie de différents projets portant sur ce thème (pour un récit plus complet, voir Hutchinson & Hayward, 2005). Ces ressources centralisées permettent une diffusion beaucoup plus large de la philosophie et des pratiques d'évaluation formative. Elles encouragent aussi le recours à des réseaux et à des experts pour faciliter le processus de mise en place de ces pratiques (la stratégie d'évaluation pour les apprentissages pose davantage de problèmes, comme nous l'avons déjà indiqué).

Le soutien de l'équipe de direction

Il est difficile d'accompagner les méthodes pédagogiques innovantes lorsque la direction de l'établissement n'est pas très coopérative (Gardner, Harlen, Hayward *et al.*, 2010). L'aspect hiérarchique du programme, du moins tel qu'il est souvent perçu, lui donne plus de crédibilité car il va au-delà de l'enthousiasme de quelques enseignants volontaires. Par ailleurs, l'octroi de crédits supplémentaires permettant de s'associer au projet pilote peut contribuer à renforcer le soutien de l'encadrement.

Le développement d'une expertise locale

La reconnaissance des expérimentations locales se traduit par l'attribution d'un statut d'expert local aux initiateurs du programme. En Angleterre, la concurrence entre établissements oblige ces derniers à se distinguer les uns des autres. Être à l'avant-garde de l'évaluation pour les apprentissages représente un atout supplémentaire, d'autant plus que d'autres dispositifs (comme par exemple autour de l'intelligence émotionnelle) demeurent moins convaincants et moins ancrés dans les travaux de recherche.

LES RISQUES D'UN SOUTIEN POLITIQUE

Si le soutien politique est nécessaire à la mise en œuvre d'une réforme, il comporte au moins trois risques. Le premier risque, qui est de voir les décideurs politiques dénaturer les intentions originales d'une innovation, a été discuté précédemment à propos du dispositif d'évaluation pour les apprentissages en Angleterre. Même si elle est bien intentionnée, l'adoption d'une expérimentation locale peut faire l'objet de malentendus et de simplifications abusives. Le deuxième risque est celui de l'imposition d'une politique par le centre : dans des systèmes éducatifs où se multiplient les projets, toute centralisation de l'expérimentation peut être perçue comme un fardeau supplémentaire pour les équipes, d'où la méfiance des enseignants et des établissements quand la mise en œuvre de cette expérimentation s'accompagne d'exigences bureaucratiques. Le dernier risque concerne la diffusion du dispositif. La manière dont les politiques centralisées sont parfois mises en œuvre empêche que la généralisation de l'expérimentation atteigne ceux pour lesquels le dispositif est pourtant une clé essentielle de leur réussite, à savoir les enseignants de terrain. En Angleterre,

la formation des enseignants à une innovation se traduit par des séances très formatées, suivies par un ou deux enseignants par établissement, qui doivent à leur tour former l'ensemble de leurs collègues. Or c'est rarement le cas du fait d'un emploi du temps extrêmement serré. Par conséquent, ces derniers ne savent guère ce que l'on attend d'eux (pour plus de détails, voir Stoll & Stobart, 2005).

CONCLUSION

Cet article porte sur le processus d'appropriation et de généralisation des expérimentations pédagogiques locales par les agences nationales. À partir d'une étude de cas au Royaume-Uni, nous avons vu les avantages et les inconvénients d'un tel processus. Dans un pays comme l'Angleterre où la culture

du résultat est forte, une expérimentation peut se trouver intégrée à un agenda politique plus complexe et être détournée de son esprit en cours de route. Néanmoins il faut saluer les progrès réalisés pour encourager les établissements à modifier leurs pratiques pédagogiques et à mobiliser de nouvelles ressources pour impliquer les équipes de direction et les enseignants. Les pratiques de détournement, la perception d'une action imposée d'en haut et la diffusion d'un dispositif à un public autre que celui initialement visé constituent aussi des risques lorsqu'une expérimentation fait l'objet d'une reprise en main centralisée. Les initiateurs de ces projets locaux doivent surtout veiller à ce que l'intention et l'esprit originel de l'expérimentation ne soient pas dilués dans la déclinaison d'une politique nationale en programmes.

Gordon Stobart
g.stobart@ioe.ac.uk
Institute of Education,
Université de Londres, Angleterre

BIBLIOGRAPHIE

- ALLAL L. & MOTTIER LOPEZ L. (2005). « L'évaluation formative de l'apprentissage : revue de publications en langue française ». In OCDE, *L'évaluation formative : pour un meilleur apprentissage dans les classes secondaires*. Paris : OCDE, p. 265-290.
- ASSESSMENT REFORM GROUP (1999). *Assessment for learning: Beyond the black box*. Cambridge : Cambridge University Press.
- ASSESSMENT REFORM GROUP (2002). *Assessment for learning: Research-based principles to guide classroom practice*. Cambridge : Cambridge University Press.
- BLACK P. & WILLIAM D. (1998). *Inside the black box: Raising standards through classroom assessment*. Londres : King's College Press.
- BLACK P. & WILLIAM D. (2003). « In praise of educational research: Formative assessment ». *British Educational Research Journal*, vol. 29, n° 5, p. 623-637.
- CLARKE S. (2001). *Unlocking formative assessment. Practical strategies for enhancing pupils' learning in the primary classroom*. Londres : Hodder & Stoughton.
- CROOKS T. (2007). « Principles for intelligent accountability, with illustrations from education » [document audio]. Dunedin : University of Otago.
- DAUGHERTY R. (2007). « Developing classroom assessments ». *American Teacher*, vol. 92, n° 4, p. 7-7.
- DEPARTMENT OF CHILDREN, SCHOOLS AND FAMILIES (2008). *The assessment for learning strategy*. Londres : Department of Children, Schools and Families.
- GARDNER J., HARLEN W., HAYWARD L. & STOBART G. (2010). *Developing teacher assessment*. Londres : Open University Press.
- GILLBORN D. & YOUDELL D. (2000). *Rationing education. Policy, practice, reform, and equity*. Buckingham : Open University Press.
- HATTIE J. & TIMPERLEY H. (2007). « The power of feedback ». *Review of Educational Research*, vol. 77, n° 3, p. 81-112.
- HUTCHINSON C. & HAYWARD L. (2005). « The journey so far: Assessment for learning in Scotland ». *Curriculum Journal*, vol. 16, n° 2, p. 225-248.
- KLUGER A. & DENISI A. (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, p. 254-284.
- O'NEILL O. (2002). *A question of trust*. Cambridge : Cambridge University Press.
- SADLER D. (1998). « Formative assessment: Revisiting the territory ». *Assessment in Education*, vol. 5, n° 1, p. 77-84.
- STOBART G. (2008). *Testing times. The uses and abuses of assessment*. Londres : Routledge.
- STOLL L. & STOBART G. (2005). « The key stage 3 strategy: What kind of reform is this? » *Cambridge Journal of Education*, vol. 35, n° 2, p. 225-238.