

Quels usages des évaluations standardisées des élèves ?

Bruno TROSSEILLE

Dans les lignes qui suivent, on n'envisage les évaluations standardisées des élèves que du point de vue de leur mise en œuvre institutionnelle à grande échelle par le ministère de l'éducation nationale. On n'évoquera pas ici les autres dispositifs d'évaluations standardisées, tels les tests utilisés par les psychologues de l'éducation ou les grilles critériées comme le LSUN ou le CECRL, ou encore les tests de compétences en langues proposés par des organismes extérieurs au MEN.

Histoire et usages des évaluations standardisées

Depuis une quarantaine d'années, le Ministère de l'Éducation nationale a mis en place, des dispositifs nationaux d'évaluation des acquis des élèves reposant sur des épreuves standardisées. Sa direction statistique (DEPP) est également maître d'œuvre pour la France de diverses évaluations internationales. Le développement des évaluations standardisées apparaît en effet comme indispensable pour rendre compte du système et le piloter. Trois grandes périodes, qui se recouvrent peu ou prou, caractérisent le développement de ces dispositifs.

Dans une première période, de la fin des années 1970 à la fin des années 1980, les dispositifs mis en œuvre à l'école et au collège en fin d'année, couvrent progressivement, niveau scolaire par niveau scolaire, l'ensemble des disciplines. Il s'agit, au regard des programmes en vigueur, d'établir un constat, d'apprécier l'état du système, de rendre compte des acquisitions des élèves aux responsables de la politique éducative du ministère. Il s'agit aussi de nourrir le débat public. Durant cette décennie, dans le cadre d'un « observatoire permanent des acquis des élèves », chaque niveau scolaire, du primaire au lycée, fera l'objet, une année donnée, d'une évaluation de type « bilan », sur des échantillons d'élèves.

Une deuxième période, à partir de 1989, est occupée par la mise en place d'évaluations diagnostiques « de masse », conséquences de la loi sur l'éducation de 1989. Le rapport qui lui est annexé relève que moins d'un élève sur deux arrive au collège avec une maîtrise suffisante de la lecture. Il précise l'urgence de la mise en œuvre d'un véritable plan sur l'apprentissage de la lecture qui « fera l'objet d'une évaluation auprès de tous les élèves entrant en CE2 et en Sixième ». Cette évaluation est un outil mis à leur disposition « ... pour déceler, de façon précise et dès le début de l'année scolaire, les difficultés de vos élèves et vous permettre, dans toute la mesure du possible d'y apporter rapidement une réponse ». Les évaluations sur échantillons du type de celles de la première décennie se font alors plus rares.

La troisième période, depuis le début des années 2000, voit se systématiser, à côté des enquêtes internationales, des évaluations sur échantillons pour un bilan des acquis des élèves en fin d'école primaire et en fin de collège (CEDRE) et pour le calcul des indicateurs de maîtrise des compétences du socle commun. Durant cette période, de nouvelles évaluations « de masse » apparaissent brièvement, ayant pour but le repérage des élèves en difficulté vis-à-vis du socle commun (objectif de mise en place de remédiations ciblées en début de CE1 notamment, de 2005 à 2007), puis d'évaluations « bilans-diagnostic » (en fin de CE1 et en milieu de CM2 de 2009 à 2012) succédant à l'arrêt des évaluations « de masse » en CE2 et en 6^e.

Rappelons que l'on peut distinguer, sur le plan conceptuel, les dispositifs « d'évaluation diagnostique » et les dispositifs « d'évaluation bilan », différant dans leurs objectifs, dans les modalités de mise en œuvre, dans l'exploitation et l'utilisation de leurs résultats :

- **axées sur les élèves**, les évaluations diagnostiques sont des outils professionnels à l'usage des enseignants. Elles doivent permettre de cerner individuellement les

compétences et les difficultés de chaque élève et d'orienter le travail de chaque élève et de la classe en fonction des résultats. Elles doivent aussi permettre de dégager des priorités de formation continue « de proximité ». En raison de leurs objectifs, de leur conception et de leur renouvellement annuel, elles ne sont pas adaptées à la comparaison dans le temps. Leurs résultats ne peuvent être comparés dans l'espace, entre les différentes classes d'une même école ou entre les différents collèges d'un département, que si les conditions de passation et de correction sont strictement respectées.

- **axées sur des populations**, les évaluations-bilans sont des outils pour le pilotage d'ensemble du système éducatif. Leur finalité, leurs objectifs et la méthodologie de leur construction ne permettent pas de rendre compte de résultats individuels pertinents pour chaque élève, chaque classe ou chaque établissement. Elles doivent permettre d'agir au niveau national sur les programmes des disciplines, sur les organisations des enseignements, sur les contextes de l'enseignement, sur des populations caractérisées. Sous certaines conditions méthodologiques, leurs résultats peuvent être comparés dans le temps.

Le ministère a souvent entretenu la **confusion entre ces deux types d'évaluation**. Ainsi, au niveau national, de 1989 à 2007, une remontée des résultats des évaluations « de masse » en CE2 et en 6^e est organisée sur des échantillons représentatifs d'élèves. Ces résultats nationaux, sont présentés comme des repères destinés à aider les enseignants à faire une analyse individuelle des freins rencontrés par leurs élèves dans les apprentissages. S'ils ne peuvent – au sens psychométrique – être comparés d'une année à l'autre, ils ont pu, ici ou là, être utilisés comme des indicateurs d'évolution des acquis des élèves, voire du système éducatif.

Cette confusion a encore été accentuée entre 2009 et 2012 avec la mise en place des évaluations « de masse » en fin de CE1 et en milieu de CM2, auxquelles le ministère assignait alors ces deux objectifs simultanés.

L'impact des résultats des grandes enquêtes nationales et internationales sur la politique éducative

Les évaluations standardisées à grande échelle, se sont multipliées depuis le début des années 2000 : au niveau national, notamment CEDRE¹, Socle commun ; au niveau international, notamment PISA (compréhension de l'écrit, maths et sciences), PIRLS (lecture en CM1), TIMSS (maths et sciences, niveaux CM1, 4^e et Terminale scientifique).

Ces programmes d'enquête vouent une attention particulière aux comparaisons temporelles, afin de pouvoir juger des progrès réalisés par les systèmes éducatifs et de les rapprocher de caractéristiques structurelles, sociales, éducatives, etc. La question de la mesure de l'évolution du niveau des élèves dans le temps est donc centrale. Dans cette perspective, il est frappant de constater la convergence des résultats des enquêtes nationales et des enquêtes internationales depuis une dizaine d'années. Au-delà des différences d'évolution selon les disciplines, on observe un accroissement des écarts entre les élèves les plus et les moins compétents. Parmi les pays de l'OCDE, la France est un des pays où la différence de scores entre les élèves issus de milieux très défavorisés et très favorisés est la plus grande. En outre, le lien entre la performance de l'élève et son statut socioéconomique et culturel est l'un des plus élevés parmi les pays de l'OCDE. Cette différenciation est observée dès le début de l'école primaire et ne fait que s'accroître au fil des années de scolarisation.

¹ Cycle des évaluations disciplinaires réalisées sur échantillons : sur un cycle de cinq ans, maîtrise de la langue française, langues étrangères, histoire-géographie, éducation civique, sciences expérimentales, mathématiques.

Si les grandes lois de programmation se réfèrent à PISA (ainsi en est-il dans le préambule de présentation du Socle commun en 2005 ou de la loi de Refondation de 2013), qu'en est-il concrètement dans les politiques mises en place par le ministère ?

Par exemple, les résultats des évaluations en compréhension de l'écrit, tant de CEDRE que de PIRLS, montrent une faiblesse de l'enseignement des stratégies de compréhension à l'école primaire. Les programmes de 2016 en ont tiré les conséquences et mettent en exergue la nécessité d'enseigner cette compréhension de manière explicite dès le cycle 2. On peut toutefois s'interroger sur les moyens donnés à la formation des enseignants pour réaliser cet objectif. La même question se pose relativement à l'enseignement des mathématiques dont les résultats aux évaluations ont dramatiquement chuté depuis une vingtaine d'années.

En revanche, la question du redoublement a récemment refait surface alors que les résultats de PISA semblent plutôt favorables aux pays qui font très peu redoubler, et malgré toutes les études nationales et internationales qui convergent pour juger le redoublement à la fois inefficace, coûteux et inéquitable, études qui vont contre l'opinion commune.

Si dans certains pays comme l'Allemagne, les résultats d'évaluations comme PISA, ont eu un impact considérable, aboutissant dès le début à une prise de conscience et à un changement dans les politiques éducatives, l'impact des résultats d'évaluations semble en France beaucoup plus mitigé. Si les diagnostics sont partagés, les conséquences en termes de politique éducative peuvent différer largement selon les orientations politiques des ministres en exercice.

Perspectives

Aujourd'hui, d'une façon générale, l'évaluation des acquis des élèves peut répondre à trois objectifs :

1. fournir aux enseignants des outils afin d'enrichir leurs pratiques pédagogiques en évaluant mieux les acquis de leurs élèves ;
2. disposer d'indicateurs permettant de mesurer, au niveau national, les performances de notre système (évolutions temporelles et comparaisons internationales) ;
3. doter les « pilotes de proximité » (recteurs, DASEN, IEN) d'indicateurs leur permettant de mieux connaître les résultats des écoles et d'effectuer une vraie régulation.

Si l'on est capable de répondre de façon adaptée aux deux premiers, le troisième a reçu jusqu'à récemment des réponses inadéquates, détournant des procédures qui ne lui correspondaient pas.

Des expériences récentes (Socle 6^e en 2016 et 2017), utilisant les technologies numériques, cherchent à répondre à ces besoins.

Bruno TROSSEILLE
Ancien chef de bureau de la DEPP