

Quels types de résultats fournissent les enquêtes de la DEPP et les évaluations nationales ?

Quels sont les effets sur les pratiques enseignantes ?

Jean-François Chesné(*)

En complément du dossier Évaluation du Bulletin Vert no 497, voici un article de Jean-François Chesné. Ces quelques pages sont la retranscription, par l'orateur lui-même, de son intervention au séminaire du 13 mars 2012 de l'APMEP.

Jean-François Chesné se proposait de répondre, sur la vingtaine de minutes qu'a duré son intervention, à deux questions, Quels types de résultats fournissent les enquêtes de la DEPP et les évaluations nationales ? Quels sont les effets sur les pratiques enseignantes ? avec pour point d'ancrage les évaluations à l'école primaire. On se souvient des passions suscitées par ces évaluations, des questions soulevées et des inquiétudes exprimées, ... Les évaluations CM2 se passent désormais en juin, soulevant d'autres interrogations sur leur exploitation : ne risquent-elles pas de devenir un « examen de passage en sixième » ? Elles feront certainement couler encore beaucoup d'encre, mais le propos développé ici mène, à travers le récit d'une expérimentation, à une réflexion sur nos pratiques.

Répondre à ces deux questions (surtout en très peu de temps) est évidemment un défi, la formulation de la première question montrant d'ailleurs l'importance d'un éclaircissement sur le sujet. J'ai donc conçu cette présentation, forcément partielle, en mettant en avant un certain nombre d'idées fortes et des éléments de langage, et en évitant les aspects théoriques, qu'ils soient statistiques ou didactiques, afin qu'elle puisse être reprise en totalité ou en partie par certains d'entre vous en formation.

Avant de répondre à ces deux questions, je vais commencer par en poser une troisième : de quelles évaluations parle-t-on ? Une première distinction importante est en effet à faire, notamment auprès des enseignants, entre l'évaluation réalisée en classe et l'évaluation à grande échelle, effectuée au niveau national ou international. Ces évaluations ne peuvent être que différentes parce qu'elles ne se font pas à la même échelle d'abord – au même « grain » diront certains – et qu'elles n'ont pas les mêmes fonctions ; elles n'ont donc pas les mêmes contenus ni les mêmes modalités en général. En effet, si l'évaluation en classe peut et doit avoir des formes très variées parce qu'elle est réalisée par un enseignant donné pour « ses » élèves, sortir des

(*) Direction de l'évaluation, de la prospective et de la performance (DEPP)
jean-francois.chesne@education.gouv.fr

moments classiques étiquetés « contrôles », s'appuyer sur l'oral, utiliser les TICE, l'évaluation à grande échelle, pour des contraintes de natures diverses, ne peut avoir qu'un format standardisé.

Je ne m'attarderai pas aujourd'hui sur l'évaluation en classe, je dirai simplement que je la conçois presque exclusivement à visée formative, en tout cas dans le champ de la scolarité obligatoire (et peut-être même au-delà). À ce sujet, je partage le point de vue de nombreux travaux anglo-saxons qui mettent en avant l'importance de la mise en œuvre effective dans la classe d'une évaluation tournée vers l'apprentissage (*assessment for learning*), voire comme modalité d'apprentissage (*assessment as learning*). Gardons enfin à l'esprit que le degré de familiarité des élèves en classe avec un type de tâche intervient dans la réussite à cette tâche dans une évaluation nationale.

Quant aux évaluations à grande échelle, après un long débat sur leur caractère diagnostique ou bilan, ce qu'il est important de comprendre aujourd'hui, c'est qu'elles peuvent être, à un niveau donné, réalisées soit sur un échantillon d'élèves, soit sur l'ensemble des élèves.

Prenons deux exemples actuels de ces deux types d'évaluations : CEDRE (Cycle d'évaluations disciplinaires réalisées sur échantillon) mené par la DEPP, et plus particulièrement par le bureau de l'évaluation des élèves, et les évaluations nationales CE1/CM2 pilotées par la Direction générale de l'enseignement scolaire (DGESCO), avec un appui statistique de la DEPP depuis 2010. Hormis le fait que ces évaluations soient placées toutes les deux à des moments clefs du cursus scolaire et élaborées en regard des objectifs fixés dans les programmes scolaires nationaux (à la différence des évaluations PISA), presque tout ce qui les caractérise est à distinguer.

- Leurs objectifs d'abord : pour l'une, l'objectif principal est d'apprécier les connaissances et les compétences des élèves dans leur ensemble et de mettre en relation les résultats obtenus avec les politiques éducatives nationales. Se déroulant tous les 6 ans pour une discipline donnée, CEDRE ne recherche pas une analyse fine des items qui sont proposés dans les tests, item par item, mais vise la production d'informations plus globales dans la perspective de comparaisons temporelles. Les évaluations CE1/CM2 sont quant à elles d'abord destinées à renseigner les enseignants sur les acquis de chaque élève, avec communication des résultats aux familles, et à aider les enseignants à réguler leur organisation pédagogique, dans et en dehors de leur classe. Un autre objectif est de rendre visibles à leurs yeux les progressions annuelles des apprentissages, telles qu'elles sont proposées dans les programmes 2008.
- Les tâches proposées aux élèves sont également différentes : beaucoup de QCM et quelques exercices demandant une production écrite pour CEDRE, non libérés, c'est-à-dire non rendus publics, alors que les évaluations CE1/CM2 demandent aux élèves essentiellement des productions écrites. Les items sont libérés par la nature même des tests.
- Les méthodologies enfin : outre la nécessité de garantir la représentativité des échantillons d'élèves, la méthodologie de CEDRE est strictement définie par

les psychométriciens du bureau chargé de l'évaluation des élèves, de la sélection des items lors de pré-tests jusqu'au traitement et à l'analyse des résultats ; cette méthodologie repose sur un système de cahiers tournants, c'est-à-dire que tous les élèves ne passent pas les mêmes tests, ce qui demande une grande rigueur dans la composition des cahiers d'évaluation, mais ce qui permet aussi de disposer de très nombreux items . C'est d'ailleurs ce système qui permet de proposer un nombre suffisant d'items identiques d'une année sur l'autre pour assurer une comparaison temporelle. Par ailleurs, les protocoles de passation des tests et de correction des cahiers sont eux aussi standardisés. Pour les évaluations CE1/CM2, si la volonté d'assurer et d'améliorer à la fois la qualité des items et la comparabilité des tests est bien réelle, le fait que les élèves passent tous le même test, que la passation et la correction soient assurées par les enseignants eux-mêmes, fournit des informations à exploiter au niveau local, mais rend délicat le calcul d'indicateurs synthétiques, et encore plus délicate une comparaison temporelle.

Nous avons donc bien deux types d'évaluations complémentaires, et donc deux types de résultats, et sans doute aussi des publics différents auxquels s'adressent ces résultats :

- CEDRE, par la construction d'une double échelle, renseigne au niveau national sur les niveaux des élèves et les caractérise par des compétences, capacités ou connaissances atteintes ou acquises. Par exemple, à l'issue de l'école primaire en 2008, CEDRE indique qu'environ 28 % des élèves avaient une bonne ou très bonne maîtrise de ce qui est attendu en fin d'école, 31 % constituaient un groupe intermédiaire, et 41 % des élèves constituaient trois groupes dits de bas ou très bas niveaux et ne possédaient que des capacités très fragiles ou très locales. Les domaines concernés sont la connaissance des grands nombres, des nombres décimaux, le calcul mental, les opérations sur les nombres décimaux et la résolution de problèmes de plus d'une étape. Dans sa conclusion, la Note d'information de la DEPP consacrée à cette enquête précise que ces trois groupes représentaient en 2008 38 % des élèves qui allaient entrer en sixième, et estime que ces élèves ne maîtrisaient pas suffisamment les notions mathématiques attendues en fin d'école primaire pour réussir de façon autonome leur scolarité au collège. A l'issue du collège en 2008, on trouvait aussi 44 % d'élèves répartis dans les trois groupes les plus faibles, avec des difficultés particulières sur les fractions, les pourcentages, les nombres relatifs, le calcul mental, l'entrée dans l'algèbre, le raisonnement déductif, et le domaine des grandeurs et mesures.
- Quant aux évaluations nationales CE1/CM2, elles fournissent des taux de réussite nationaux des élèves à certains exercices et renseignent localement les enseignants sur ceux que leurs élèves savent faire ou ne pas faire. Ces renseignements sont donc très intéressants pour eux-mêmes, individuellement, au niveau de la classe, de l'école, ou même d'une circonscription, mais il faut avoir conscience de la variabilité des taux de

réussite selon la tâche. Par exemple, deux multiplications posées sur des entiers ont été données en 2010 et 2011 : l'une, 39×57 , (Item 79, 2010) faisait intervenir les tables de 7 et de 9 et fut réussie par 51,6 % des élèves alors que l'année suivante, l'autre, 14×35 (Item 83, 2011), qui ne convoquait que la connaissance des tables de 3, 4 et 5 relevant du CE1, fut réussie par 73,9 % des élèves. Un autre exemple est la différence des taux de réussite à deux exercices proposés en 2011, relevant de la proportionnalité, et qui peuvent apparaître comme très voisins. **Le directeur doit acheter des cahiers et des livres pour l'école. 6 livres coûtent 150 €. Combien coûtent 9 livres ?** (Item 88, 2011, 51,5 %), et **10 objets identiques coûtent 22 €. Combien coûtent 15 de ces objets ?** (Item 89, 2011, 30,7 %) Dans les deux cas, le rapport de linéarité est 1,5 (c'est-à-dire que la quantité à calculer est égale à la somme de la quantité initiale et de sa moitié), mais une stratégie correspondant au passage par l'unité fait intervenir dans le deuxième cas un nombre décimal. Sur ces deux exemples, on voit tout le travail passionnant d'exploration et d'aide qu'un enseignant peut et doit faire auprès de ses élèves, mais aussi tout le travail d'analyse, du côté de la tâche et du côté des élèves, qui est à sa charge.

En conclusion intermédiaire, je dirai que « naturellement » la tentation est grande quand on dispose des résultats obtenus par chaque élève à un test de chercher à les agréger pour en faire des indicateurs départementaux, académiques ou nationaux. Mais, pour rendre compte d'évolutions dans le temps, cela nécessite des précautions méthodologiques très strictes, dont le revers de la médaille est que ces précautions imposent des contraintes sur l'ensemble du processus qui restreignent une utilisation à portée immédiate des tests et de leurs résultats par les enseignants. Il est donc naturel de s'interroger sur la diffusion et l'utilisation des résultats de ces deux types d'évaluation, des publics auxquels ils s'adressent, et sur leurs effets sur les pratiques enseignantes.

Pour répondre sur ce dernier sujet, et pour prolonger la première partie de mon propos, je vais vous présenter dans ses grandes lignes une expérimentation menée actuellement par mon bureau à la DEPP. Le Projet pour l'Acquisition de Compétences par les Élèves en Mathématiques (PACEM) est mis en œuvre depuis septembre 2010 en CM1/CM2 à Marseille et en 6^e/5^e dans l'académie de Créteil. Il s'agit d'aller au-delà des simples constats locaux et de la communication des taux de réussite des élèves à leurs familles, et de montrer comment l'appropriation et l'utilisation d'une évaluation standardisée par des enseignants peut influencer sur la modification de leurs pratiques et contribuer à l'amélioration des acquis des élèves.

À travers une collaboration avec les IEN, les conseillers pédagogiques, le groupe de pilotage départemental pour les mathématiques et les IA-IPR, le dispositif repose sur la mise en œuvre, à partir des contenus et des résultats d'un pré-test, d'une formation courte de certains enseignants, fondée sur l'exploration d'une « Zone proximale de développement des pratiques ». Cette formation est destinée à la fois à modifier des pratiques individuelles ordinaires et à encourager un travail collectif au sein des

établissements. Très brièvement, cela consiste à partir des représentations et des pratiques des enseignants pour « légitimer » le test et ses utilisations, c'est-à-dire pour faire en sorte que les enseignants s'en approprient les contenus, puis à en analyser les résultats afin qu'ils les intègrent à la fois dans l'organisation de leur propre enseignement et dans une réflexion au niveau de l'école ou du collège. Une des spécificités de la formation est d'associer les résultats du pré-test à la fois à ceux de tests antérieurs réalisés ou non sur échantillon et à des résultats issus de la recherche en didactique des mathématiques, afin de mettre en relief certains attendus des programmes. L'impact de l'expérimentation sur les acquis des élèves est évalué par une comparaison avec les résultats à un deuxième test proposé en fin d'année scolaire.

Je suis tout à fait conscient qu'un tel projet, par les nombreuses variables qu'il met en jeu, nécessite d'être prudent dans la présentation de ses résultats. Toutefois, à l'issue de la première année, l'augmentation des scores des élèves expérimentateurs par rapport à ceux d'élèves témoins est très significative en CM1, notamment pour les élèves faibles, et significative en 6^e avec une variabilité importante selon les collèges. À ce stade du dispositif, j'aurais donc envie de formuler l'hypothèse prudente qu'une « certaine didactique appliquée » s'inscrit plus facilement dans les pratiques des professeurs des écoles que dans celle des professeurs de mathématiques du second degré, comme si les premiers avaient plus de marges pour aménager les conditions de mise en activité des élèves dans la classe et les rendre efficaces dès lors qu'ils ont accès à un minimum d'enjeux didactiques, alors que les contraintes qui pèsent sur les seconds, notamment dans l'organisation du temps de classe, du point de vue des tâches et des scénarios associés, rendraient plus difficile, pour résumer, le passage de l'enseignement à l'apprentissage.

En conclusion, dans un contexte où toutes les enquêtes de la DEPP montrent une augmentation du nombre d'élèves de bas niveaux et mettent en évidence l'impact de leur environnement socio-économique sur leur réussite, je voudrais attirer votre attention sur une partie d'un processus dialectique que je crois que l'on considère à tort qu'il va de soi. En effet, la mise en œuvre du socle commun, à l'école et peut-être encore plus au collège, s'est faite par une entrée par la validation des compétences. Les enseignants ont donc été et sont actuellement confrontés à de réelles difficultés relatives à l'évaluation, les modalités de prise d'information, de communication, ou la détermination de seuils de décision à partir desquels un élève maîtriserait ou ne maîtriserait pas une compétence. Or, les enseignants sont habitués à produire un système d'évaluation des performances de leurs élèves, lui-même déterminé par un ensemble complexe comprenant les représentations du métier, des mathématiques et de leur enseignement, les conditions d'exercice, les ressources disponibles au quotidien. On voit donc bien qu'il ne peut y avoir d'évolution des acquis des élèves que par une évolution des tâches qui leur sont proposées, et des déroulements qui leurs sont associés, pendant le temps de la classe, en amont, dans un contexte d'apprentissage. Mais cela demande que les enseignants disposent d'outils disciplinaires, intellectuels et pratiques, qui leur permettent de mettre en œuvre cette évolution, ce qui nécessite une réflexion approfondie dans plusieurs

directions, à laquelle la DEPP participe par l'évolution des méthodologies et des instruments d'évaluation qu'elle propose.

Références

- IGEN, *L'enseignement des mathématiques au cycle 3 de l'école primaire*, 2006.
- VANDEBROUCK F. (coordinateur), *La classe de mathématiques : activités des élèves et pratiques des enseignants*. Toulouse, Octares Éditions, 2007.
- PASTOR J.-M. et BRUN A., *Les compétences en mathématiques des élèves en fin d'école primaire*, NI 10.17, 2010.
- CHESNÉ J.-F., *Les acquis des élèves en calcul à l'issue de l'école*, Éducation & formations, N°79, 2010
- Huguet T. et Brun A., *Les compétences en mathématiques des élèves en fin de collège*, NI 10.18, 2010.
- HCE, *Les indicateurs relatifs aux acquis des élèves*, 2011.
- CHESNÉ J-F et PROST S., *PACEM : une expérimentation sur l'utilisation d'évaluations standardisées des acquis des élèves par les enseignants*, Éducation & formations, N° 81, 2012.