

En 2007, des indicateurs portant sur la maîtrise des compétences de base requises en fin d'école et en fin de collège, en français et en mathématiques ont été élaborés. La mise au point de ces indicateurs a suivi une méthodologie spécifique qui confronte les attentes et les exigences du système éducatif avec les performances observées des élèves. Selon les niveaux scolaires et les disciplines, 10 % à 20 % des élèves ne maîtrisent pas les compétences de base. Ces proportions sont cohérentes avec celles observées dans les évaluations nationales et internationales.

Méthodologie de l'évaluation des compétences de base en français et en mathématiques en fin d'école et en fin de collège

Le contexte

Parmi les indicateurs de résultats retenus dans le cadre de la LOLF – loi organique relative aux lois de finances – figurent les proportions d'élèves qui maîtrisent les compétences de base en français et en mathématiques en fin d'école et en fin de collège (voir *Indicateurs de résultats de la LOLF*, p. 6).

Plus précisément, ces quatre proportions doivent être déclinées selon les catégories suivantes : le secteur public, le secteur de l'éducation prioritaire, le réseau « ambition réussite » et le secteur privé. À l'origine, ces seize taux devaient être calculés dans chacune des trente académies. Mais, pour des raisons logistiques et financières, en 2007, les résultats sont valables uniquement au plan national.

Ces indicateurs répondent à une exigence forte : la comparabilité dans le temps. En effet, ils doivent permettre de mesurer l'évolution des acquis des élèves – en termes de compétences de base – et l'atteinte des objectifs fixés. Ils seront calculés chaque année à partir des résultats obtenus par un échantillon d'élèves à des tests standardisés. Ces tests sont donc complètement différents et indépendants, dans leur mise en œuvre et leur objectif, des évaluations diagnostiques qui sont des outils pédagogiques.

Pour répondre à cette demande, la Direction de l'évaluation, de la prospective et de la performance a conçu un dispositif d'évaluation *ad hoc* qui a été mis en place pour la première fois en 2007, après deux années d'expérimentation en 2005 et 2006. Guidé par les enseignements des évaluations menées aux niveaux national et international, ce dispositif a suivi une méthodologie précise que détaille cette *Note d'Information*.

Les contraintes

L'élaboration de ce dispositif tient compte de deux contraintes principales : le temps et le coût. Les résultats doivent être disponibles chaque année avant la fin de l'été pour la préparation du débat budgétaire au Parlement. Pour cette raison, les tests sont passés en mars.

À cette contrainte de temps, s'ajoute la contrainte financière qui implique la construction d'un test sous forme de QCM (questions à choix multiples) : les élèves choisissent la bonne réponse parmi celles qui sont proposées. Ce format de questions assure une correction rapide, fiable et économique. En contrepartie, une interrogation sous cette forme exclut, d'emblée, l'évaluation de certaines compétences comme, par exemple, la production écrite (voir p. 2 le paragraphe « Définition des compétences de base »).

La démarche

La première étape du travail a consisté, en collaboration avec les services pédagogiques du ministère et les inspections générales, à établir une *définition des compétences de base* pour chaque discipline (français et mathématiques) et pour chaque niveau (CM2 et troisième). Sur cette base, des groupes de concepteurs, composés d'inspecteurs et d'enseignants, ont participé à l'*élaboration des épreuves*. Un nombre important de questions a été mis au point et a fait l'objet d'une *expérimentation* en 2006. Les données issues de cette expérimentation ont été mises en regard avec les jugements (attentes, niveaux d'exigence) des groupes de concepteurs, selon une méthodologie spécifique. Cette confrontation a conduit à la *détermination du seuil* à partir duquel un élève peut être considéré comme maîtrisant les compétences de base. Le *test de 2007* s'est appuyé sur une sélection d'items (questions) issus de l'expérimentation et a concerné des échantillons d'élèves représentatifs au plan national. Les *premiers résultats* sont présentés ici. Afin de garantir la pertinence des comparaisons, ces résultats sont accompagnés d'intervalles de confiance qui prennent en compte les erreurs d'échantillonnage inhérentes à ce type d'enquête.

Définition des compétences de base

Les compétences de base visées par l'indicateur font référence au socle commun de connaissances et de compétences (voir *Indicateurs de résultats de la LOLF*, p. 6). Au moment de l'élaboration des tests, en octobre 2005, si la notion de socle commun existait bien, aucun texte officiel ne définissait encore précisément son contenu. Celui-ci ayant été connu en juillet 2006, il en a été tenu compte pour consolider l'ensemble des items.

L'Inspection générale de l'éducation nationale (IGEN) et le service en charge de l'élaboration des programmes scolaires (DGESCO) ont mis au point des documents de travail dans lesquels ont uniquement

été retenues les compétences dites « de base », en français et en mathématiques, en fin d'école et en fin de collège.

Ces compétences ont ainsi été définies à partir d'éléments issus des programmes scolaires et en relation étroite avec le socle commun de connaissances et de compétences. Mais les compétences évaluées sont celles que la forme d'évaluation retenue – en l'occurrence un test standardisé composé de QCM – permet de mesurer. De ce fait, en français, les compétences liées à l'expression orale et celles liées à l'expression écrite ne sont pas prises en compte. De la même manière, en mathématiques, l'écriture effective de nombres, la construction de figures géométriques ou encore la production de démonstrations ont été écartées.

Élaboration des épreuves

Afin d'élaborer les items destinés à évaluer les compétences retenues, des groupes de travail ont été réunis dans chaque discipline, à chaque niveau : ils étaient composés d'enseignants, de conseillers pédagogiques, d'inspecteurs de l'éducation nationale (IEN, IA-IPR, IGEN) et de représentants du service ministériel chargé

des programmes. Ces groupes ont mis au point un grand nombre d'items afin de permettre le renouvellement et l'évolution des tests dans les prochaines années.

En 2005, une première enquête-pilote avait été conduite dans deux académies (Besançon et Bordeaux) à partir d'items élaborés en collaboration avec l'université de Rennes 2. Les items mis au point en 2006 se sont inspirés, pour une partie, de ces tests, notamment en mathématiques.

Mathématiques

En CM2, sur les 244 items mis au point en 2006 par le groupe de travail, 221 ont été retenus pour l'expérimentation. Ces items ont été répartis dans treize « blocs » de 17 items, en respectant au maximum le poids de chaque champ évalué (géométrie, grandeurs et mesure, etc.). Ces treize blocs ont été agencés de manière à constituer treize cahiers différents, chacun composé de quatre blocs (voir l'encadré ci-dessous). Chaque élève de l'échantillon a ainsi travaillé pendant une heure sur un cahier constitué de 68 items (17 × 4).

Un schéma similaire a été adopté en troisième : sur les 294 items mis au point, 169 ont été retenus pour l'expérimentation 2006. Ils se répartissent en treize blocs de

Les « cahiers tournants » pour expérimenter un grand nombre d'items

La méthode des « cahiers tournants » est utilisée pour évaluer un nombre important d'items sans allonger le temps de passation. Elle consiste à répartir les items dans des cahiers différents qui comportent des items communs. Pour cela, les items sont d'abord répartis dans des « blocs ». Un bloc d'items est tout simplement un ensemble d'items, un ou plusieurs exercices par exemple. Les blocs doivent avoir la même unité de temps de passation, chacun prévu pour une demi-heure par exemple. Les cahiers sont ensuite composés de trois ou quatre blocs d'items.

Afin de relier les résultats obtenus à tous les items, la répartition des blocs dans les cahiers doit répondre à certaines contraintes. Par exemple, il est préférable que chaque association (chaque paire) de blocs soit testée afin de pouvoir calculer toutes les intercorrélations entre items. Il est aussi possible d'ajouter une contrainte de position : par exemple, chaque bloc doit apparaître une fois au début du cahier, une autre fois en milieu de cahier, etc.

Ces contraintes restreignent les possibilités de combinaison. Dans le cadre de ce test, la

configuration retenue est celle de treize blocs d'un quart d'heure répartis dans treize cahiers contenant chacun quatre blocs. Dans cette configuration, chaque bloc apparaît dans quatre cahiers différents, à quatre places différentes. De plus, toutes les associations de blocs sont testées.

Les blocs s'agencent de la manière suivante dans les cahiers :

N° du cahier	Place dans le cahier			
	1	2	3	4
cahier 1	B1	B3	B4	B8
cahier 2	B2	B4	B5	B9
cahier 3	B3	B5	B6	B10
cahier 4	B4	B6	B7	B11
cahier 5	B5	B7	B8	B12
cahier 6	B6	B8	B9	B13
cahier 7	B7	B9	B10	B1
cahier 8	B8	B10	B11	B2
cahier 9	B9	B11	B12	B3
cahier 10	B10	B12	B13	B4
cahier 11	B11	B13	B1	B5
cahier 12	B12	B1	B2	B6
cahier 13	B13	B2	B3	B7

treize items et chacun des treize cahiers comporte donc 52 items (13 × 4). Au total, plus de trois heures de test ont été expérimentées, à l'école ainsi qu'au collège.

Français

En français, la situation est différente car les items peuvent rarement être « décontextualisés ». Ils se rapportent à un stimulus, le plus souvent un texte écrit. De ce fait, il est moins aisé de construire, comme en mathématiques, des blocs équilibrés pour les « cahiers tournants ». La démarche envisagée repose sur la construction d'épreuves « parallèles ». Ainsi, en CM2, deux versions d'un même test ont été élaborées sur la base des épreuves de l'enquête-pilote de 2005. Chacune des versions comprend des exercices de lecture de consignes, des items d'orthographe, un travail à partir d'un dictionnaire de langue rare et des questions de compréhension et de maîtrise de la langue sur deux textes narratifs et un texte documentaire. La première version est composée de 67 items et la seconde de 75 items. Deux cahiers ont alors été constitués pour tenir compte des effets liés à l'ordre de passation : dans le premier cahier la version 1 est suivie de la version 2, le second cahier suit l'ordre inverse.

En troisième, cinq « corpus » ont été mis au point en 2006 pour évaluer les compétences de base en français. Un « corpus » est un ensemble de 66 à 122 items qui ont en commun un thème général, qui respecte le poids des compétences visées et dont la durée de passation n'excède pas une heure. Pour l'expérimentation 2006, deux « corpus » ont principalement été évalués. Afin de relier les résultats, quatre cahiers d'une heure et demie ont été constitués à partir des items d'un corpus et de la moitié des items de l'autre corpus.

Expérimentation 2006

Pour l'expérimentation 2006, dans la perspective éventuelle de décliner les indicateurs au niveau académique, les trente académies ont pris part à l'opération. Ainsi, les activités liées au dispositif d'évaluation ont été réparties entre la DEPP et les services académiques. Pour simplifier les procédures, la moitié des académies a

expérimenté les tests de français et l'autre moitié les tests de mathématiques. De plus, la taille des échantillons était volontairement réduite : environ 500 élèves de CM2 et 500 élèves de troisième par académie (c'est-à-dire 10 à 26 écoles et 15 collèges par académie). Le relais des services académiques au niveau local et la mise au point de plates-formes de communications efficaces ont permis de recueillir des données de qualité (taux de réponses proches de 100 %, respect des consignes de passation, etc.). Ce dispositif « allégé » qui concernait 15 000 élèves par niveau scolaire (CM2 et troisième) a également permis d'évaluer la charge de travail et le coût financier qu'impliquerait la déclinaison de ces indicateurs au niveau académique.

S'agissant de l'exploitation des résultats aux tests, les réponses des élèves ont fait l'objet d'analyses statistiques préalables. Ainsi, selon les tests, de 10 % à 25 % des items se sont avérés faiblement discriminants, c'est-à-dire peu liés à la dimension générale visée, le plus souvent en raison d'ambiguïtés dans le questionnement ou dans les réponses proposées. Par ailleurs, aucun effet de fatigue ou d'entraînement n'a pu être décelé. De plus, les éventuels comportements de réponses au hasard, dont on pouvait craindre l'existence à cause du format des items, ne sont pas apparus à l'analyse. Enfin, les réponses des élèves se structurent de manière unidimensionnelle : autrement dit, les items sont fortement corrélés entre eux et à la dimension générale mesurée dans chacune des disciplines, à chaque niveau.

Détermination du seuil de maîtrise

L'expérimentation a permis de recueillir les résultats des élèves à un ensemble d'items qui mesurent l'acquisition des compétences de base, telles qu'elles ont été définies en amont. Cependant, ces résultats ne permettent pas de calculer directement la proportion des élèves qui maîtrisent ces compétences de base, car le « degré de maîtrise » n'a pas encore été défini. En effet, les items peuvent être de difficulté très variable, quand bien même ils portent sur une compétence dite de « base ». Par exemple, en mathématiques, il est possible de rendre

très complexes des items dont la résolution ne fait appel qu'aux opérations les plus élémentaires. De ce fait, pour être considéré comme maîtrisant les compétences de base, un élève doit-il réussir toutes les questions qui lui sont proposées ? les trois quarts ? la moitié ? C'est ce seuil qui doit être fixé, seuil à partir duquel nous considérerons que les élèves maîtrisent les compétences de base.

La détermination de ce seuil ne s'impose pas d'elle-même. Plusieurs possibilités peuvent être envisagées.

La première consiste à le fixer relativement à la distribution des élèves. Par exemple, pour les évaluations annuelles de la DEPP sur échantillons, le seuil du groupe 1 est intentionnellement déterminé de manière à identifier les 15 % des élèves les plus faibles. Dans un second temps, les compétences de ces élèves sont décrites en fonction des items qu'ils réussissent ou échouent. Il s'agit-là d'une définition relative comme celle, par exemple, qui fixe le seuil de pauvreté à la moitié du revenu médian. Cette approche n'est pas celle retenue ici car elle n'intègre pas le niveau requis *a priori* permettant de considérer qu'un élève maîtrise les compétences de base.

À l'opposé, une deuxième approche consiste à fixer dès le départ une règle de calcul qui, à partir des résultats observés, indique si l'élève maîtrise ou non les compétences de base. Ce choix est hasardeux car il suppose que la réussite des élèves au test soit effectivement conforme à celle prévue au départ. Or, les décalages entre les attentes et la réalisation sont nombreux dans le domaine de l'évaluation des compétences, ne serait-ce que pour des raisons liées à l'erreur de mesure (mauvaise formulation des questions, ambiguïtés dans les choix de réponses proposés, etc.).

Le point de vue envisagé ici a consisté à combiner ces deux approches antagonistes en confrontant les données statistiques issues de l'expérimentation avec les attentes et les exigences des groupes de travail. Dans un premier temps, ces experts n'ont pas eu accès aux résultats statistiques et ils ont été conviés à un travail individuel en suivant deux méthodes qui s'inspirent de la littérature scientifique disponible dans ce domaine (voir l'encadré p. 4).

Deux méthodes pour déterminer le seuil de maîtrise des compétences de base

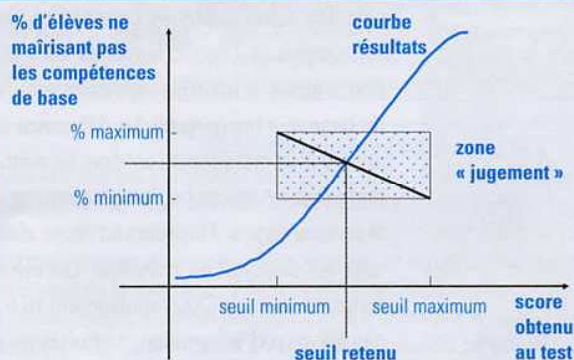
– Première méthode

Pour chaque discipline – français et mathématiques – et pour chaque niveau – CM2 et troisième – les experts ont répondu à la question suivante : « Sur la base de votre expérience professionnelle, quels sont, selon vous, le pourcentage minimum et le pourcentage maximum d'élèves ne maîtrisant pas les compétences de base ? » (Le point de vue adopté est celui de la non-maîtrise des compétences de base, plutôt que de la maîtrise ; ce point de vue s'avère plus révélateur pour identifier un seuil).

Il leur a été demandé ensuite de déterminer quel est le score – en termes de nombre de bonnes réponses – en deçà duquel les élèves peuvent être considérés comme ne maîtrisant pas les compétences de base. Plus précisément, il leur a été demandé de déterminer un score minimal acceptable correspondant à une notation « large » et un score maximal acceptable correspondant à une notation « sévère ». Un score acceptable leur a été présenté de la manière suivante : « un score qui correspond à votre propre perception et qui vous semble légitime du point de vue de ce que l'institution scolaire peut attendre d'un élève. »

Les experts ont travaillé à partir des items du test en suivant cette consigne : « La borne inférieure relève d'un niveau d'exigence minimal. En pratique, il s'agit de déterminer les questions fondamentales qui doivent absolument être réussies. A contrario, la borne supérieure renvoie à l'idée que les élèves devraient réussir toutes les questions posées, si on s'en tenait à une vision rigide des compétences de base. En pratique, pour déterminer cette borne supérieure, il s'agit simplement d'éliminer certaines questions qui ne semblent pas relever stricto sensu des compétences de base. »

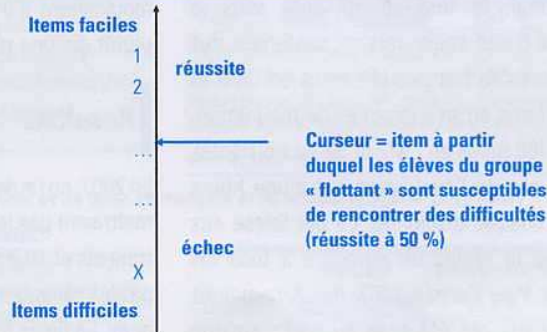
À chaque expert correspond alors une zone de jugement (cf. figure ci-dessous). Cette zone témoigne à la fois des attentes et du niveau d'exigence de chaque expert. À partir des données issues de l'expérimentation, il est possible de calculer pour chaque score le pourcentage d'élèves situés en deçà de ce score et de tracer la courbe correspondante. Le point d'intersection entre cette courbe et la diagonale de la zone de jugement fournit le score-seuil retenu.



Source : MEN-DEPP

– Deuxième méthode

L'objectif final est de produire un pourcentage d'élèves qui maîtrisent les compétences de base. Ainsi, de fait, deux groupes d'élèves sont distingués : ceux qui maîtrisent et ceux qui ne maîtrisent pas ces compétences. Évidemment, la réalité n'est pas aussi tranchée et les groupes ne sont pas aussi clairement distincts l'un de l'autre. Certains élèves sont susceptibles d'être classés dans l'un ou l'autre des deux groupes selon des paramètres extérieurs comme par exemple la fatigue, l'environnement, les conditions de passation, etc. Ces élèves forment un groupe dit « flottant ».



Source : MEN-DEPP

Dès lors, il s'agit de repérer un niveau de difficulté d'items permettant de caractériser le mieux ce groupe d'élèves « flottant ». C'est ce niveau de difficulté qui nous permettra de fixer le seuil de maîtrise des compétences de base.

À partir des résultats issus de l'expérimentation, les items ont été classés par champ et par difficulté croissante. Les items du début de chaque liste ont été les mieux réussis lors de l'expérimentation et ceux de la fin les moins bien réussis (cf. figure ci-dessus). Pour chaque champ, les élèves du groupe « flottant » sont susceptibles de réussir la majorité des items du début de la liste et d'échouer la majorité des items de la fin de la liste.

Il est important de noter que la liste des items n'était accompagnée d'aucun résultat complémentaire, d'aucun chiffre ou taux de réussite qui aurait pu influencer le choix des experts.

Il a été demandé aux experts de repérer dans la liste fournie la position de l'item à partir duquel peut s'opérer « la bascule », c'est-à-dire à partir duquel les élèves du groupe « flottant » vont commencer à rencontrer des difficultés. Pour permettre une certaine souplesse, il était possible d'indiquer trois items consécutifs plutôt qu'un seul.

Références

La première approche s'inspire de la méthode d'Hofstee (cf. Norcini, 2003) et la seconde de celle d'Angoff (cf. Laveault et Grégoire, 2002).

Laveault, D. et Grégoire, J. (2002). *Introduction aux théories des tests en psychologie et en sciences de l'éducation* (2^{ème} édition). Bruxelles : DeBoeck-Université.

Norcini, J. J. (2003). Setting standards on educational tests, *Medical Education*, 37, 464–469.

La première méthode compare directement les jugements et les attentes des experts avec les données statistiques. La seconde méthode s'appuie sur les compétences attendues d'un groupe virtuel d'élèves dit « flottant », c'est-à-dire un groupe d'élèves fragiles, situés juste autour du seuil de maîtrise des compétences de base.

Les résultats de ces deux méthodes ont abouti à l'établissement de seuils de maîtrise des compétences de base, différents selon les experts et selon la méthode. Ces résultats ont donc fait l'objet d'une discussion plénière qui a conduit à un consensus autour d'un seuil.

En plaçant sur la même échelle les élèves et les items, il est alors possible de

décrire les compétences maîtrisées par les élèves situés au-dessus du seuil établi (graphique 1). Cette description permet d'affiner et de valider ce seuil.

Test 2007

Le test doit fournir une mesure précise du niveau de compétences des élèves situés

autour du seuil déterminé à l'étape précédente. En effet, il n'est pas nécessaire que le test soit précis sur les élèves les plus compétents. Par conséquent, le choix des items tient compte de leur difficulté, de manière à ce qu'ils soient adaptés au niveau de compétences correspondant au seuil retenu. La sélection des items pour le test de 2007 a également respecté l'équilibre entre les champs, les sous-compétences visées, les supports, etc. Au final, le test se présente sous la forme d'une seule version contenant des épreuves de français (75 items en CM2 et 127 items en troisième) et de mathématiques (68 items en CM2 et 52 en troisième). La durée de l'évaluation est d'une heure pour chaque discipline, ce qui laisse aux élèves le temps de répondre à tous les items. Pour l'année 2007, des échantillons nationaux ont été tirés au sort : environ 2 000 élèves par zone (secteur public hors éducation prioritaire, le réseau de réussite scolaire, le réseau ambition réussite et le secteur privé) et par niveau scolaire (CM2 et

troisième). Au total, près de 8 000 élèves de CM2 et 8 000 élèves de troisième ont passé les tests. Comme toute enquête procédant par échantillonnage, les résultats doivent être accompagnés d'une marge d'incertitude, que l'on peut estimer en fonction des données. En l'occurrence, les intervalles de confiance varient de $\pm 1,8\%$ à $\pm 3,8\%$ selon le niveau scolaire et la zone de scolarisation. Ces restrictions ne permettront probablement pas de déceler de mouvement d'une année sur l'autre mais plutôt sur une plus longue période.

Résultats

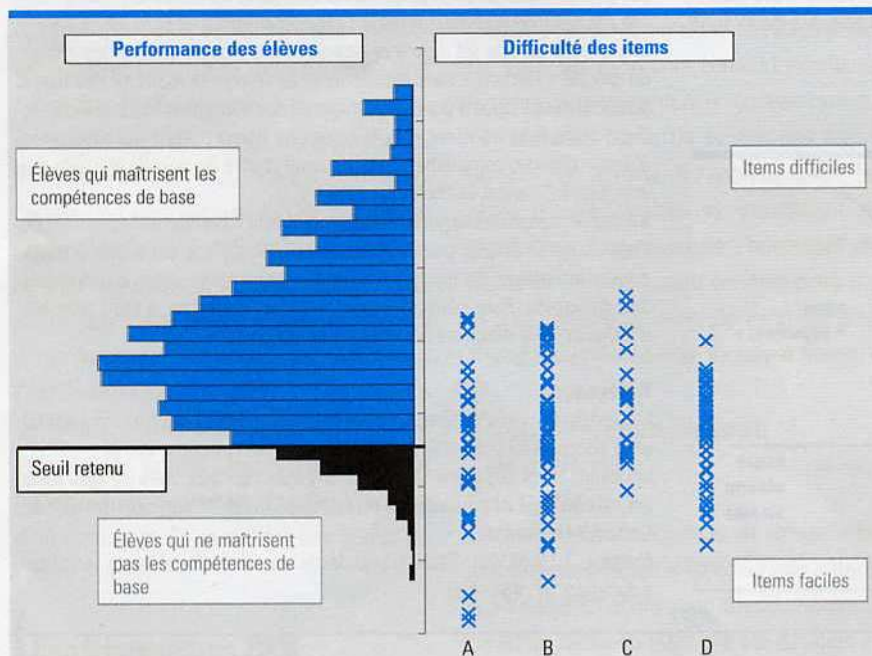
En 2007, en fin de CM2, 13,7 % des élèves ne maîtrisent pas les compétences de base en français et 10,2 % des élèves ne maîtrisent pas les compétences de base en mathématiques. En fin de troisième, 20,1 % des élèves ne maîtrisent pas les compétences de base en français et 10,6 % des élèves ne maîtrisent pas les compétences de base en mathématiques (voir les graphiques 1 et 2 qui

présentent les indicateurs de la LOLF, dont la définition est renversée : il s'agit des proportions d'élèves qui maîtrisent les compétences de base).

Ces proportions sont du même ordre que celles observées dans d'autres dispositifs d'évaluations. Ainsi, parmi les indicateurs retenus dans le cadre du processus de Lisbonne engagé par les pays de l'Union européenne, figure la proportion d'élèves de 15 ans ayant de faibles résultats en lecture, c'est-à-dire situés dans les groupes 0 et 1 de l'échelle de compréhension l'écrit de PISA (Programme international pour le suivi des acquis des élèves). En 2006, en France, cette proportion s'élevait à 21,8 % (voir *Pour en savoir plus*, p. 6). Au niveau national, les tests de la JAPD (Journée d'appel de préparation à la défense), passés par les jeunes français d'environ 17 ans, ont pour objectif de repérer les faibles lecteurs. En 2006, le pourcentage de jeunes en difficulté de lecture s'élevait à 11,7 %, auxquels s'ajoutent 9,6 % de lecteurs médiocres pour lesquels la lecture reste une activité laborieuse (voir *Pour en savoir plus*, p. 6). Ces résultats, s'ils proviennent d'évaluations dont les objectifs et les populations visées diffèrent, sont néanmoins cohérents avec ceux observés ici en français en fin de troisième.

Les indicateurs de la LOLF sont également calculés selon quelques caractéristiques. En français, à l'école, les filles sont plus nombreuses à maîtriser les compétences de base que les garçons. La différence est encore plus marquée au collège. En mathématiques, en revanche, les garçons ont un léger avantage sur les filles à l'école, avantage qui disparaît en troisième. Les élèves en retard en fin de CM2 représentent 16,5 % des élèves de l'échantillon ; en troisième, ce sont 35,6 % des élèves qui ont au moins une année de retard. Que ce soit à l'école ou au collège, la proportion d'élèves qui ne maîtrisent pas les compétences de base, en français ou en mathématiques, est nettement plus élevée parmi les élèves en retard que parmi les élèves « à l'heure ». Ce constat ne suffit pas à condamner le redoublement mais il renvoie à des études qui montrent son inefficacité (voir *Pour en savoir plus*, p. 6). Enfin, les résultats selon la « zone » de scolarisation sont très contrastés, en particulier concernant l'éducation

GRAPHIQUE 1 – Une même échelle pour les élèves et les items



Lecture : les items et les élèves ont été positionnés sur la même échelle, à partir des résultats observés lors de l'expérimentation, grâce aux modèles de réponse à l'item. Dans la partie gauche est représentée la distribution des élèves selon leur score, c'est-à-dire le nombre d'élèves (longueur de la barre en abscisse) en fonction du score obtenu (intervalles de scores en ordonnée). Dans la partie droite, chaque croix représente un item, qui est positionné en fonction de sa difficulté, du mieux réussi au moins bien réussi. Les items les plus faciles sont placés dans le bas du graphique, en face des élèves les moins performants et, parallèlement, les items les plus difficiles font face aux élèves les plus performants. Plus précisément, chaque item est positionné à un niveau tel que les élèves situés à ce niveau ont un taux de réussite de 50 % à cet item. Ainsi, les élèves situés au-dessus du seuil retenu réussissent à plus de 50 % les items placés en dessous de ce seuil. Il est donc possible de décrire les items que maîtrisent ces élèves. De plus, les items sont distingués selon le champ auquel ils appartiennent, par exemple géométrie, nombres, etc. (ici A, B, C, D). Cela permet d'affiner la description des tâches que maîtrisent les élèves situés au-delà du seuil.

Source : MEN-DEPP

GRAPHIQUE 2 – Proportion d'élèves de fin de CM2 qui maîtrisent les compétences de base en français et en mathématiques (LOLF 2007)

En français, environ 86 % des élèves de fin de CM2 sont capables :

Lecture	de chercher des informations en se référant à l'organisation d'un dictionnaire ; de comprendre globalement un texte littéraire ou documentaire court et d'y prélever des informations ponctuelles explicites ;
Maîtrise des outils de la langue	de maîtriser partiellement l'automatisation de la correspondance grapho-phonologique ; d'identifier les principaux temps de l'indicatif pour les verbes les plus fréquents ; de reconnaître les règles les plus simples d'orthographe lexicale et grammaticale.

En mathématiques, environ 90 % des élèves de fin de CM2 sont capables :

Exploitation de données numériques	de prélever une information dans un tableau ; de résoudre des problèmes simples relevant de l'addition et de la soustraction ;
Connaissance des nombres et calcul	de passer d'une écriture en lettres à une écriture en chiffres (ou le contraire) et de comparer, d'additionner et de soustraire des nombres entiers naturels ; de reconnaître le double ou la moitié d'un nombre entier « familier » ; de passer d'une écriture en lettres à une écriture sous forme fractionnaire (ou le contraire) de fractions simples ;
Espace et géométrie	de reconnaître visuellement un triangle, un triangle rectangle, un rectangle, un carré ; de reconnaître par une représentation en perspective un cube ou un parallélépipède rectangle ;
Grandeurs et mesure	de mesurer la longueur d'un segment ; d'utiliser les unités de mesure des durées (sans calculs).

	Français	Mathématiques
Ensemble	86,3 %	90,6 %
Garçons	83,7 %	91,1 %
Filles	89,0 %	88,1 %
À l'heure	91,5 %	94,2 %
En retard	60,2 %	68,0 %
Public	86,8 %	90,6 %
ZEP	77,7 %	83,8 %
RAR*	72,0 %	75,4 %
Privé	92,1 %	93,1 %

* RAR : réseau ambition réussite
Lecture : 86,3 % des élèves de CM2 maîtrisent les compétences de base en français. L'intervalle de confiance (I-I) de cet indicateur est de $\pm 2,1$ %.

Source : MEN-DEPP

GRAPHIQUE 3 – Proportion d'élèves de fin de troisième qui maîtrisent les compétences de base en français et en mathématiques (LOLF 2007)

En français, environ 80 % des élèves de fin de troisième sont capables :

Compréhension des textes	de reconnaître un texte explicatif ; de distinguer les principaux genres de textes ; de prélever des informations explicites ; de faire des inférences simples ; et de donner une interprétation d'un texte sans difficulté de compréhension, à partir d'éléments simples ;
Maîtrise des outils de la langue	d'identifier les structures syntaxiques fondamentales ; d'analyser les principales formes verbales ; de faire un emploi pertinent du vocabulaire courant ; de repérer différents niveaux de langue ; de reconnaître les règles d'orthographe et de ponctuation, d'usage courant.

En mathématiques, environ 90 % des élèves de fin de troisième sont capables :

Organisation et gestion de données, fonctions	d'utiliser une représentation graphique dans des cas simples (lecture des coordonnées d'un point, lien avec un tableau numérique dans une situation de proportionnalité, détermination des données d'une série statistique) ; de calculer la moyenne d'une série statistique ; de traiter des problèmes simples de pourcentages ;
Nombres et calculs	de comparer des nombres décimaux relatifs écrits sous forme décimale ; d'utiliser les opérations élémentaires dans une situation concrète ;
Grandeurs et mesures	d'effectuer pour des grandeurs (durée, longueur, contenance) un changement d'unités de mesure (h min en min, km en m, L en cL) ; de calculer le périmètre d'un triangle dont les longueurs des côtés sont données ; de calculer l'aire d'un carré, d'un rectangle dont les longueurs des côtés sont données dans la même unité ;
Géométrie	d'identifier des figures simples à partir d'une figure codée et d'en utiliser les éléments caractéristiques (triangle équilatéral, cercle, rectangle) ; d'écrire et d'utiliser le théorème de Thalès dans un cas simple ; de reconnaître un patron de cube ou de parallélépipède rectangle.

	Français	Mathématiques
Ensemble	79,9 %	89,4 %
Garçons	74,5 %	89,4 %
Filles	85,5 %	89,4 %
À l'heure	90,7 %	96,4 %
En retard	60,3 %	76,8 %
Public	80,3 %	90,5 %
ZEP	70,5 %	81,2 %
RAR*	54,8 %	71,0 %
Privé	87,7 %	93,2 %

* RAR : réseau ambition réussite
Lecture : 89,4 % des élèves de 3^{ème} maîtrisent les compétences de base en mathématiques. L'intervalle de confiance (I-I) de cet indicateur est de $\pm 1,7$ %.

Source : MEN-DEPP

Indicateurs de résultats de la LOLF

Dans le cadre des indicateurs de résultats de la LOLF, les « proportions d'élèves maîtrisant les compétences de base (définies en référence au socle commun) en français et en mathématiques » s'inscrivent dans l'objectif, pour le premier degré, de « conduire tous les élèves à la maîtrise des compétences de base exigibles au terme de la scolarité primaire » et, pour le second degré, de « conduire le maximum d'élèves aux niveaux de compétences attendues en fin de scolarité et à l'obtention des diplômes correspondants », et ce dans le secteur public comme dans le secteur privé. De plus, dans l'objectif d'« accroître la réussite scolaire des élèves en zones difficiles et des élèves à besoins éducatifs particuliers », ces indicateurs doivent être également calculés pour les écoles – resp. les collèges – en éducation prioritaire (EP) et les écoles – resp. les collèges – en réseau ambition réussite (RAR).

Théoriquement, ces indicateurs devraient être déclinés au niveau académique. Mais si, au niveau académique, pour chaque indicateur chiffré, un minimum de 1 000 élèves est requis, cela signifie 4 000 élèves dans chacune des trente académies, et ce en fin de CM2 et en fin de troisième, c'est-à-dire 120 000 élèves par niveau scolaire, ce qui implique des moyens logistiques et financiers considérables. Or, l'expérimentation de 2006 a montré la faible efficacité d'un tel dispositif. En effet, cela nécessite un investissement important de la part des services académiques pour un apport limité : étant basés sur un échantillon, les résultats ne peuvent être retournés ni aux élèves ni aux établissements. En outre, l'indicateur académique est affecté d'une marge d'erreur qui ne permettra pas d'observer d'évolutions significatives d'une année sur l'autre. Dans cette optique, un intervalle de cinq ou six années serait plus instructif.

prioritaire : près de la moitié des élèves du « réseau ambition réussite » ne maîtrisent pas les compétences de base en français au collège.

Thierry Rocher, Jean-François Chesné et Sylvie Fumel, DEPP B2

Pour en savoir plus

« L'évolution des acquis des élèves de 15 ans en culture mathématique et en compréhension de l'écrit. Premiers résultats de l'évaluation internationale PISA 2006. », *Note d'Information* 08.08, MEN-DEPP, janvier 2008.

« Les évaluations en lecture dans le cadre de la journée d'appel et de préparation à la défense – Année 2006 », *Note d'Information* n° 07.25, MEN-DEPP, mai 2007.

Le redoublement au cours de la scolarité obligatoire : nouvelles analyses, mêmes constats, Les Dossiers, n° 166, MEN-DEP, mai 2005.