



L'orientation scolaire et professionnelle

36/2 | 2007
Varia

Évaluation des comportements et des compétences scolaires par les enseignants et prédiction des performances et des parcours à l'école élémentaire et au collège

Teacher rating of behaviours and competences, and prediction of academic performances and of later school pathways in primary and secondary school

Philippe Guimard, Olivier Cosnefroy et Agnès Florin



Édition électronique

URL : <http://journals.openedition.org/osp/1372>
DOI : 10.4000/osp.1372
ISSN : 2104-3795

Éditeur

Institut national d'étude du travail et d'orientation professionnelle (INETOP)

Édition imprimée

Date de publication : 15 juin 2007
Pagination : 179-202
ISSN : 0249-6739

Référence électronique

Philippe Guimard, Olivier Cosnefroy et Agnès Florin, « Évaluation des comportements et des compétences scolaires par les enseignants et prédiction des performances et des parcours à l'école élémentaire et au collège », *L'orientation scolaire et professionnelle* [En ligne], 36/2 | 2007, mis en ligne le 01 juin 2010, consulté le 19 avril 2019. URL : <http://journals.openedition.org/osp/1372> ; DOI : 10.4000/osp.1372

Ce document a été généré automatiquement le 19 avril 2019.

© Tous droits réservés

Évaluation des comportements et des compétences scolaires par les enseignants et prédiction des performances et des parcours à l'école élémentaire et au collège

Teacher rating of behaviours and competences, and prediction of academic performances and of later school pathways in primary and secondary school

Philippe Guimard, Olivier Cosnefroy et Agnès Florin

Introduction

- 1 Mieux comprendre les parcours des élèves constitue un enjeu important pour définir des politiques éducatives centrées sur la réussite scolaire à l'école élémentaire. Depuis plusieurs décennies, le ministère de l'Éducation nationale met régulièrement en place des suivis sur plusieurs années de panels d'élèves afin d'identifier les parcours scolaires et de déterminer les facteurs impliqués dans la réussite et l'échec à l'école. Parallèlement, depuis 1989, des évaluations diagnostiques standardisées sont réalisées à l'entrée en CE2 et en sixième afin de permettre aux enseignants d'identifier les acquis et les difficultés de leurs élèves et d'adapter leurs pratiques pédagogiques aux besoins réels de chacun d'entre eux. Dans ces différents dispositifs, l'évaluation des compétences des élèves dans des domaines fondamentaux comme la maîtrise de l'oral, de l'écrit ou des mathématiques occupe une place privilégiée.
- 2 Toutefois, pour mieux comprendre les parcours scolaires des élèves, il peut s'avérer utile de ne pas limiter la prise des informations au seul domaine cognitif, mais de disposer également d'une évaluation des comportements et des compétences des élèves en classe au moyen d'évaluations réalisées par les enseignants. Divers auteurs considèrent en effet

que les enseignants sont dans une position optimale pour décrire des savoir-faire et des difficultés comportementales impliquées dans l'adaptation scolaire des élèves (Chevrie-Müller, 1999 ; O'Neill & Liljequist, 2002 ; Mercer, Algozzine & Trifiletti, 1988 ; Taylor, Anselmo, Foreman, Schatschneider & Angelopoulos, 2000 ; Zazzo, 1978). Le fait que les enseignants soient au contact quotidien des élèves les conduit à développer une expérience particulièrement riche pour repérer une diversité de comportements et de savoir-faire chez leurs élèves et pour en percevoir l'évolution (McCloskey, 1990). Rapide et peu onéreux, ce type d'évaluation semble ainsi plus dynamique que les évaluations cognitives standardisées et plus proches de la réalité du fonctionnement de l'élève en classe (Coleman & Dover, 1993).

- 3 Pour autant, il convient de s'interroger sur la valeur des évaluations réalisées par les enseignants pour comprendre les parcours des élèves, alors même que les évaluations cognitives permettent au moins dans certains domaines d'analyser les compétences des élèves et de pronostiquer des difficultés ultérieures. En d'autres termes, les évaluations réalisées par les enseignants permettent-elles de décrire le fonctionnement scolaire des élèves et de prédire leur évolution ? Dans l'affirmative, apportent-elles des informations nouvelles par rapport à celles provenant d'évaluations standardisées ?

Évaluation des comportements et des compétences scolaires par les enseignants et prédiction des performances scolaires ultérieures

- 4 De nombreuses recherches récentes ont analysé la pertinence descriptive et prédictive de questionnaires renseignés par les enseignants dont l'objectif est de décrire les comportements que mobilisent leurs élèves en classe dans les situations d'apprentissage. Par exemple, McDermott, Leigh et Perry (2002) et McClelland et Morisson (2003) montrent que dès l'âge de 3-4 ans, les élèves se différencient du point de vue de leurs capacités d'autonomie, de coopération et d'indépendance. En grande section de maternelle, la capacité des élèves à se conformer à la demande de l'enseignant et à suivre ses directives est liée à l'adaptation des enfants à ce niveau scolaire (Foulks & Morrow, 1989). Ladd, Birsh et Buhs (1999) indiquent également qu'à ce niveau, la participation des enfants en classe et leur capacité à coopérer sont de bons prédicteurs des performances ultérieures à court terme. De même, dans une étude réalisée de la 1^{re} à la 4^e année primaire, Alexander, Entwisle et Dauber (1993) montrent que l'intérêt des enfants pour les activités scolaires, leur participation et leur attention en classe sont liés aux performances scolaires (voir, pour la motivation, Oehler-Stinette & Boykin, 2001). De plus, la valeur prédictive de ces évaluations comportementales reste sensible même après contrôle du QI et des variables sociodémographiques (McClelland, Morisson & Holmes, 2000). On peut ainsi supposer que les difficultés d'adaptation scolaire commenceraient à être repérées par les enseignants de maternelle ou de début d'école élémentaire dans les comportements des élèves avant de se traduire ultérieurement dans les performances scolaires dans les différentes matières enseignées. Comme le notent DiLalla, Marcus et Wright-Phillips (2004), ces travaux suggèrent également qu'en intervenant directement sur certains comportements en classe, il serait possible d'aider les enfants dans leur processus d'apprentissage, d'augmenter leurs performances académiques et ainsi d'éviter que des difficultés se cristallisent et deviennent résistantes aux interventions.

- 5 D'autres études visent à déterminer la valeur prédictive des compétences scolaires évaluées par les enseignants sur les performances académiques ultérieures. Ainsi, Quay et Steele (1998) montrent qu'une évaluation par les enseignants des performances académiques globales des élèves est prédictive de leurs performances ultérieures évaluées par une batterie d'épreuves standardisées. Meisels, Bickel, Nicholson, Xue et Atkins-Burnett (2001) obtiennent des résultats similaires à partir d'une évaluation des compétences des élèves en mathématiques, en langage oral et dans le domaine de l'écrit réalisée au moyen du questionnaire *Work Sampling System* (WSS) renseigné par les enseignants. De plus, les résultats de cette recherche montrent que cette évaluation est prédictive des performances ultérieures indépendamment du sexe et de facteurs sociodémographiques. Dans une étude réalisée de la 1^{re} année à la 3^e année primaire, Hecht et Greenfield (2002) visent à déterminer les raisons pour lesquelles les enseignants sont capables de prédire de manière fiable les compétences futures de leurs élèves en lecture. Leurs résultats montrent que des compétences liées à la lecture et évaluées par l'enseignant sont de meilleurs prédicteurs que certains comportements de l'enfant en classe (l'enfant suit les règles de la classe ; il s'investit dans les tâches scolaires).
- 6 De façon générale, un large consensus semble ainsi se dégager de la recherche sur le fait, d'une part, que les meilleurs prédicteurs comportementaux se réfèrent à des comportements centrés sur les apprentissages (McClelland, Morisson & Holmes, 2000 ; McClelland & Morisson, 2003) et, d'autre part, que dès l'école maternelle, les évaluations comportementales et/ou scolaires effectuées par les enseignants présentent une bonne valeur prédictive des performances et des difficultés scolaires ultérieures (Belsito, Ryan & Brophy, 2005 ; Cadieux & Boudreault, 2002 ; Crosby & French, 2002 ; Florin, 1991 ; Florin, Guimard & Nocus, 2002 ; Guimard & Florin, 2001 ; Guimard, Florin & Nocus, 2002 ; Teisl, Mazzocco & Myers, 2001).

Comparaison de la valeur prédictive des évaluations réalisées par les enseignants et des évaluations cognitives standardisées

- 7 L'examen de la littérature montre qu'à défaut d'avoir un grand degré de précision, les évaluations des enseignants sont significativement corrélées avec des évaluations de type psychologique (cognitives, langagières, par exemple) réalisées dans la même période (Agostin & Bain, 1997 ; Hoge, 1983 ; Hoge & Coladarci, 1989 ; Kenny & Chekaluk, 1993 ; Meljac, Kugler & Mogenet, 2001 ; Quay & Steele, 1998 ; Zazzo, 1978). Compte-tenu de ces résultats, une des questions importantes est alors de savoir quelle est la valeur prédictive des évaluations réalisées par les enseignants par rapport à des évaluations cognitives standardisées.
- 8 Plusieurs études (Écalte, 1999 ; Mantzicopoulos & Morrison, 1994 ; Quay & Steele, 1998) indiquent que les évaluations par les enseignants en fin de maternelle ou en début de première année primaire prédisent mieux les performances académiques ultérieures ou le parcours scolaire des élèves que les évaluations cognitives. Toutefois, Fletcher et Satz (1984) considèrent que la valeur prédictive des évaluations réalisées par les enseignants est moins pertinente, surtout lorsqu'il s'agit de prédire les difficultés scolaires ultérieures et notamment en lecture. D'autres auteurs proposent de considérer que les deux types d'évaluation sont complémentaires. Ainsi, dans une recherche réalisée de la grande section de maternelle (*Kindergarten*) à la 3^e année primaire, Flynn et Rahbar (1998) comparent la valeur prédictive d'une évaluation par les enseignants à l'aide de la *Teacher*

Rating Scale (TSR) évaluant des domaines de compétences associées à l'acquisition de la lecture à une batterie standardisée comportant des épreuves prédictives de la compétence lexicale. Les résultats indiquent que la valeur prédictive de la TSR est moins bonne que celle de la batterie prédictive, mais que la prédiction est améliorée si l'on tient compte des deux évaluations. De même, Yen, Konold et Mc Dermott (2004) réalisent une étude auprès d'un échantillon de 1 304 élèves âgés de 6 à 17 ans afin de déterminer si une évaluation comportementale effectuée par les enseignants au moyen de la *Learning Behavior Scale* (LBS) (Mc Dermott, 1999) contribue à l'explication des performances académiques des élèves indépendamment d'une évaluation cognitive standardisée. Des analyses en pistes causales montrent que les évaluations cognitives sont les meilleurs prédicteurs des performances scolaires mais que l'évaluation des comportements réalisée par les enseignants ajoute une part modeste mais significative à l'explication des performances académiques des élèves.

Objectifs de l'étude

- 9 Les recherches visant à déterminer l'intérêt des évaluations des enseignants pour décrire et prédire la réussite scolaire varient considérablement selon les méthodologies utilisées. Les prédictions peuvent être globales ou individuelles, porter sur des enfants tout venant ou en difficulté, et être réalisées à plus ou moins long terme. Par ailleurs, le type d'évaluations sollicitées « à l'entrée » et les compétences des élèves ciblées « en sortie » varient également d'une recherche à l'autre. À l'entrée, les évaluations peuvent concerner le comportement des élèves et/ou leurs performances scolaires. Il peut également s'agir d'un pronostic de difficultés ultérieures. En sortie, les évaluations standardisées sont également variables selon que l'on évalue les performances en lecture, en mathématiques ou un niveau scolaire plus global associant plusieurs types de compétence (lecture, écriture, mathématique, par exemple).
- 10 En dépit de ces variations, l'examen de la littérature atteste que les évaluations des comportements et des compétences scolaires par les enseignants ont, en tant que telles, une bonne valeur prédictive des performances scolaires ultérieures. Ceci explique que, dans les pays anglo-saxons tout au moins, ce type d'évaluation soit assez fréquemment utilisé dans une perspective de prévention des difficultés d'apprentissage afin de repérer les enfants « à risque » et de mettre en œuvre les aides adaptées à leurs difficultés. De ce point de vue, de nombreuses recherches confirment la bonne valeur prédictive différentielle de certaines échelles d'évaluation renseignées par les enseignants (Cadioux & Boudreault, 2002 ; Fletcher, Tannock & Bishop, 2001 ; Taylor *et al.*, 2000 ; Quay & Steele, 1998). Toutefois, les résultats des recherches visant à comparer la valeur prédictive de ces évaluations par rapport aux évaluations cognitives semblent moins consensuels. À l'instar en particulier de la recherche de Yen *et al.* (2004), il semble que les évaluations cognitives standardisées prédisent mieux la réussite académique que celles des enseignants, ces dernières contribuant néanmoins à la prédiction, indépendamment du poids des évaluations cognitives et des facteurs sociodémographiques.
- 11 La présente recherche vise, d'une part, à savoir si une évaluation des comportements et des compétences des élèves en classe effectuée par les enseignants apporte un gain d'information par rapport aux évaluations cognitives standardisées dans la prédiction des performances scolaires ultérieures. Nous cherchons ainsi à confirmer les résultats de l'étude menée par Yen *et al.* (2004) en nous appuyant toutefois sur des données

longitudinales obtenues auprès d'un échantillon de 5 549 élèves suivis de la première année primaire à la première année du collège. La dimension longitudinale paraît en effet plus à même de rendre compte de la valeur écologique des résultats obtenus. Il s'agit, d'autre part, de comparer la capacité des évaluations réalisées par les enseignants et des évaluations cognitives initiales à prédire les trajectoires d'échec des élèves au cours de leur scolarité trois et six ans plus tard. Situé dans le prolongement de la recherche de Yen *et al.* (2004), cet objectif questionne l'utilité de ces deux types d'évaluation pour la prévention des difficultés d'apprentissage.

Méthode

Participants

- 12 Les données concernent 5 549 élèves suivis longitudinalement de la première année primaire à l'entrée en sixième (1^{re} année du collège) pour lesquels on dispose de l'ensemble des informations (sociodémographiques, compétences et trajectoires scolaires) pour la période considérée. Cet échantillon est issu d'un panel d'élèves scolarisés au cours préparatoire à la rentrée scolaire 1997-1998 dans une école publique ou privée sous contrat de France métropolitaine. À l'entrée en première année primaire, en 1997, ils sont âgés en moyenne de 78 mois (écart-type : 4 mois).
- 13 Plusieurs prises d'informations sur les caractéristiques des élèves, leurs compétences et leur devenir scolaire ont été réalisées en 1^{re} année d'école primaire (cours préparatoire), en 3^e année d'école primaire (CE2) et à l'entrée au collège (sixième). Les caractéristiques sociodémographiques ont été obtenues par une enquête réalisée auprès des directeurs d'école et des familles. Les variables sociodémographiques retenues sont celles qui sont les plus fréquemment associées à la réussite ou l'échec scolaire : la catégorie socioprofessionnelle des parents (origine favorisée, origine plutôt favorisée, origine « moyenne », origine défavorisée¹), le sexe, le trimestre de naissance, la nationalité (français *versus* étranger) et l'âge d'entrée à l'école maternelle (2 ans, 3 ans, 4 ans et plus) (Caille, 2001 ; Crahay, 2003 ; Duru-Bellat, 2003 ; Ferrier, 2003 ; Florin, Cosnefroy & Guimard, 2004 ; Jeantheau & Murat, 1998). Le tableau 1 représente la répartition des élèves selon les 5 variables sociodémographiques sélectionnées. Cette répartition est assez conforme à celle observée au niveau national (MEN, 2005). Cependant nous constatons une sous-représentation des classes défavorisées et de la nationalité étrangère due probablement au fait que les élèves de ces deux catégories présentent un risque élevé d'être en échec scolaire (Goldstein & Rabash, 2003) et d'être orientés vers d'autres structures.
- 14 Les informations relatives au parcours scolaire des élèves de l'échantillon sont les suivantes : en 1997, l'ensemble des élèves est entré en CP ; trois ans plus tard, en 2000, c'est-à-dire en CE2, ils sont 4,5 % à présenter un an de retard dont l'origine est un redoublement soit en CP, soit en CE1. Six ans plus tard, en 2003, ils sont 9 % à présenter un retard dont la cause est un redoublement dans les années précédentes. Ces taux sont inférieurs de 8,5 points en CE2 et de 11 points en sixième par rapport au niveau national (MEN, 2005). En accord avec les données de la recherche récente (Caille, 2004 ; Cosnefroy & Rocher, 2004 ; Crahay, 2004), le retard scolaire consécutif au redoublement est considéré dans cette étude comme un indicateur de difficultés scolaires.

Tableau 1/Table 1

Sexe	Garçon 2 788 50,2	Fille 2 761 49,8		
Nationalité	Français 5 533 96,1	Étranger 216 3,9		
Préscolarisation	2 ans 279 5	3 ans 3 510 63,2	4 ans et plus 1 760 31,7	
Origine sociale	Favorisée 1 247 22,5	Plutôt favorisée 1 207 21,8	Moyenne 1 918 34,6	Défavorisée 1 177 21,2
Trimestre de naissance	Janvier/mars 1 358 24,5	Avril/juin 1 456 26,2	Juillet/sept. 1 388 25	Octobre/déc. 1 347 24,3

Caractéristiques sociodémographiques (nombre et %) des élèves de l'échantillon (N = 5 549)
Socio-demographic characteristics (number and %) of the pupils of the sample (N = 5 549)

Mesures

- 15 Au début de la première année primaire (première quinzaine d'octobre 1997), une évaluation du comportement des élèves en classe a été renseignée par les enseignants à l'aide d'une grille d'observation se présentant sous la forme d'un questionnaire pré-établi. Les performances scolaires des élèves ont par ailleurs été évaluées collectivement par leurs enseignants selon un protocole standardisé défini par la Direction de l'évaluation et de la prospective (actuellement DEPP²). En début de 3^e année primaire puis en début de sixième (septembre-octobre 1999), tous les élèves ont été évalués en français et en mathématiques dans le cadre de la procédure d'évaluation collective définie au plan national à chacun de ces deux niveaux scolaires.
- 16 *Présentation de la grille d'observation renseignée par les enseignants (1^{re} année primaire).* La grille utilisée est une adaptation d'un outil élaboré par Florin (1991) dans le cadre du suivi longitudinal de plusieurs cohortes d'enfants afin d'examiner les liens entre leurs comportements en maternelle et leurs trajectoires scolaires à l'école élémentaire. Elle regroupe des dimensions comportementales centrées sur les apprentissages qui sont les plus prédictives en fin de maternelle/début de primaire des difficultés et des réussites scolaires ultérieures, comme l'avait montré Zazzo (1978) et comme nous l'avons confirmé dans plusieurs études ultérieures (Guimard & Florin, 2001 ; Guimard, Florin & Nocus, 2002 ; Florin, Guimard & Nocus, 2002).
- 17 La grille comporte 13 questions³ relatives aux aspects suivants : la confiance en soi (items 1 et 2), l'attention (item 3), la fatigue (item 9), l'intégration dans la classe (items 4 et 13), la rapidité (item 5), l'efficacité (item 6), la capacité d'organisation dans l'exécution

d'une tâche (item 12), l'autonomie (item 7), la maîtrise des gestes (item 8) et la participation à la conversation scolaire (items 10 et 11). Pour chaque item, les enseignants doivent répondre sur une échelle en 4 points : 1 pour un comportement jamais remarqué ou une compétence non acquise, 4 pour un comportement habituellement remarqué ou une compétence acquise, 2 et 3 pour des positions intermédiaires nuanciant le jugement entre les deux extrêmes.

- 18 Une analyse factorielle en composantes principales avec rotation oblique a été effectuée sur cet échantillon afin de déterminer la structure des items (voir tableau 2).

Tableau 2/Table 2

Items du questionnaire	Facteur 1	Facteur 2
Confiance en soi (item 1)	.80	
Confiance en soi (item 2)	.12	.87
Attention (item 3)	.77	
Intégration en classe (item 4)	.76	
Rapidité (item 5)	.81	
Efficacité (item 6)	.86	
Autonomie (item 7)	.87	
Maîtrise des gestes (item 8)	.74	
Fatigue (item 9)	.14	.56
Conversation scolaire (item 10)	.74	
Conversation scolaire (item 11)	.80	
Organisation (item 12)	.86	
Intégration en classe (item 13)	.62	

Saturation des 13 items du questionnaire renseigné par les enseignants en 1^{re} année d'école élémentaire sur les deux composantes extraites de l'analyse factorielle en composantes principales (ACP)

Saturation of the 13 items of the teacher rating scale in 1st Grade on the two components extracted from the principal components factorial analysis (PCA)

- 19 Les résultats indiquent que 11 items saturent un premier facteur représentant 62,9 % de la variance. Les items 2 et 9 semblent corrélés à un second facteur, très peu lié au premier (corrélations de .13). Les valeurs propres de ces deux facteurs sont respectivement de 6,9 et .80. Ces deux items ont été écartés de la solution finale dans la mesure où la valeur propre du facteur correspondant est inférieure à 1 et très largement inférieure à celle du premier facteur. La fidélité du questionnaire constitué de ces 11 items (alpha de

Cronbach) est de .90. Cette solution en un facteur apparaît comme satisfaisante et autorise la construction d'un score global (CS) obtenu en effectuant la somme des scores aux 11 items. Ce score donne une estimation des comportements scolaires perçus par l'enseignant.

- 20 *Évaluation des performances scolaires des élèves en début de première année primaire.*
L'évaluation des performances scolaires regroupe 5 tâches issues d'un ensemble initial de 12 épreuves (MEN, 2002). Trois épreuves (attention partagée, comparaison sociale, culture technique et connaissances de l'environnement technique) ont été écartées du fait que leur construction et/ou le codage des réponses proposées ne s'adaptait pas à des analyses quantitatives avancées. Sur les 9 évaluations restantes, quatre (épreuves de connaissances générales et connaissance de l'écrit, lecture, concepts liés au temps, compréhension orale) ont également été retirées en raison de leur assez faible consistance interne ($\alpha < .70$).
- 21 Les 5 épreuves retenues représentent un total de 74 items évaluant les dimensions suivantes : le domaine numérique (complètement de suites numériques, comparaison de collections, résolutions de problèmes arithmétiques et dénombrement ; 13 items), l'écriture (de lettres et de mots ; 23 items), les concepts liés au temps (compréhension des concepts « début/fin », « avant/après », « commence/fini », etc. ; 14 items), les compétences en prélecture (lecture de lettres, de mots, de pseudo-mots et connaissance du son des lettres ; 15 items) et la reconnaissance de nombres et de figures géométriques (identifier parmi des cibles des nombres et des figures géométriques dictés oralement ; 9 items). Pour chaque épreuve, un score a été calculé en effectuant la somme des réussites aux items correspondants (1 point par item réussi).
- 22 Les coefficients de consistance interne des cinq épreuves (alpha de Cronbach) sont présentés de manière brute et de manière corrigée (correction de Spearman Brown) (Laveault & Grégoire, 1997⁴). Le tableau 3 indique que chacune d'elles atteint un niveau d'homogénéité satisfaisant (α calculé proche ou supérieur à .70). Lorsque l'on prend en compte le nombre d'items, on constate que l'homogénéité des 5 épreuves (alpha corrigé) est similaire. Un score global de performances scolaires (PS) a donc été déterminé en calculant la somme des items composant ces 5 évaluations. La fidélité de ce score est élevée ($\alpha = .90$). Cette mesure regroupe une diversité de compétences dont on sait qu'elles sont impliquées dans la construction des apprentissages scolaires fondamentaux et fortement prédictives des performances scolaires ultérieures (Florin, Cosnefroy & Guimard, 2004 ; Jeantheau & Murat, 1998 ; Le Bastard & Suchaut, 2000 ; Mingat, 1984 ; Mingat & Richard, 1991).

Tableau 3/Table 3

	n	alpha calculé	alpha corrigé
Épreuve numérique	13	.70	.75
Concepts liés au temps	14	.68	.69
Écriture	23	.80	.72
Compétences de prélecture	15	.77	.77

Nombres et figures géométriques	9	.62	.73
---------------------------------	---	-----	-----

Consistance interne (alpha de Cronbach) des cinq épreuves retenues dans le calcul du score global de performances scolaires au cours préparatoire

Internal consistency (Cronbach' alpha) of the five tests retained in the calculation of the school performances total score in 1st Grade

- 23 *Évaluations des performances en français et en mathématiques à l'entrée au CE2 (3^e année primaire).* Dans la version proposée en 1999 pour les élèves non-redoublants (MEN, 2000), l'évaluation des compétences en français comporte 91 items répartis en 3 champs : la compréhension orale et écrite, la maîtrise des outils de la langue et la production d'écrits. L'évaluation en mathématiques est composée de 80 items appréhendant les dimensions suivantes : géométrie, mesures, compétences numériques et résolution de problèmes numériques. Un point est attribué par item réussi.
- 24 Toutefois, pour intégrer dans les analyses les performances des élèves ayant redoublé entre la 1^{re} et la 3^e année primaire, un score spécifique en français et en mathématiques a été constitué en calculant la somme des items strictement identiques à ceux de l'évaluation nationale de 1999 et de 2000 (année au début de laquelle les redoublants sont arrivés en CE2). Ce score, qui représente plus de 50 % des items de l'évaluation de 1999, permet ainsi de ramener les résultats de l'ensemble des élèves redoublants et non redoublants sur une même échelle de mesure. Les indices d'homogénéité sont élevés ($\alpha = .84$ pour le français et $\alpha = .82$ pour les mathématiques).
- 25 *Évaluations des performances en français et en mathématiques à l'entrée en sixième.* Le protocole d'évaluation nationale à l'entrée en sixième a été proposé en septembre 2002 pour les élèves non-redoublants et en septembre 2003 pour les redoublants. La version de 2002 (MEN, 2003) comporte 87 items en français et 87 items en mathématiques. Les grands domaines de compétences appréhendés en français sont analogues à ceux proposés en CE2 (compréhension orale et écrite, maîtrise des outils de la langue et production écrite). En mathématiques, les domaines évalués concernent la géométrie et les mesures, la numération et l'écriture des nombres, les techniques opératoires, les problèmes numériques et la mise en relation d'informations numériques. Le fait que les évaluations proposées en 2002 aux élèves à l'heure et en 2003 aux élèves redoublants possèdent plus de 50 % d'items communs a permis, comme précédemment, de construire en français et en mathématiques, une échelle de mesure identique pour les deux années. La consistance interne est de .78 pour l'échelle de français et de .80 pour les mathématiques.
- 26 *Récapitulatif.* Pour harmoniser les mesures, les scores globaux aux épreuves évaluant les performances scolaires en CP (PS), en CE2 et en sixième et celui de la grille d'observation (CS) ont été ramenés à un total sur 100. Le tableau 4 présente les principales caractéristiques des mesures utilisées dans cette recherche.

Tableau 4/Table 4

	Moyenne	Écart-type	Mini	Maxi	Nombre d'items	Alpha
Score global (CS) en début de CP	67,7	24,2	0	100	11	.90

Score global (PS) en début de CP	69,2	15,1	10,8	100	74	.90
Score en français en CE2	70,3	14,6	2,9	100	68	.84
Score en mathématiques en CE2	68,9	18,3	1,7	100	59	.82
Score en français en 6 ^e	75,2	15,3	0	100	38	.78
Score en mathématiques en 6 ^e	68,7	15,5	0	100	51	.80

Caractéristiques des mesures utilisées en CP, en CE2 et en sixième (moyenne, écart-type, minimum, maximum, nombre d'items et indice de consistance interne)

Characteristics of the measurements used in 1st, 3rd and 6th Grade (mean value, standard deviation, minimum, maximum, number of items and internal consistency coefficient)

Méthode d'analyse des données

- 27 Des analyses de régression ont été réalisées pour interroger le pouvoir explicatif de nos deux prédicteurs (CS et PS). Toutes les variables continues ont été normalisées afin d'apprécier leurs contributions relatives dans les équations de régression. Les modèles de régression employés sont de deux types. Lorsque nous nous intéressons à l'explication des performances scolaires (en français et en mathématiques en CE2 et en sixième) par nos deux prédicteurs, quatre régressions linéaires sont employées. Plus précisément, nous utilisons le modèle linéaire général (procédure GLM avec SAS version 9.1). Nous nous appuyons sur la statistique F pour interroger la significativité des coefficients de régression. Lorsque nous nous intéressons à l'explication du parcours scolaire (la prédiction des parcours scolaires jusqu'en CE2 et sixième) par nos deux prédicteurs, nous modélisons la probabilité d'accès en CE2 et en sixième avec redoublement grâce à des régressions logistiques (procédure LOGISTIC avec SAS 9.1). Afin de tester la stabilité des contributions des deux variables prédictives, nous intégrons les 5 variables sociodémographiques présentées précédemment dans nos modélisations de la performance et du parcours scolaire, ceci pour les deux matières et les deux niveaux examinés.
- 28 Afin d'examiner l'importance du pouvoir explicatif du score comportemental CS selon la performance ou le parcours scolaire en CE2 et en sixième, les coefficients de régression CS et PS ont été comparés. Autrement dit, nous avons examiné l'importance du pouvoir explicatif de CS, comparativement à celui de PS. Nous neutralisons ainsi les effets des métriques grâce à la comparaison des coefficients. Pour ce faire, nous avons calculé, pour chaque modèle, un ratio entre les coefficients de régression de ces deux variables : β_{PS} / β_{CS} . Le calcul de ces ratios et la comparaison de ces derniers, quelle que soit la modélisation, sont licites dans la mesure où toutes les variables continues sont standardisées. Enfin, nous avons choisi une technique de ré-échantillonnage, et plus particulièrement de *bootstrapping*, pour tenter d'inférer nos résultats à la population. Il s'agit ici de considérer l'échantillon comme une réplique de la population dont il est extrait. À partir de l'échantillon initial (N = 5549), l'objectif a consisté à extraire un nombre n de sous-échantillons avec remise (ici, $n = 1\ 000$) de taille identique (N = 5549). Nous avons utilisé la procédure SURVEYSELECT sous SAS 9.1.

Résultats

Comparaison de la valeur prédictive globale de la grille d'observation et des performances scolaires de début CP avec les performances scolaires en CE2 et en sixième

- 29 Le tableau 5 propose un extrait de la matrice de corrélation (rho de Spearman) correspondant au croisement entre respectivement, le score global à la grille d'observation (CS) et le score global de performances scolaires (PS) de début de CP, et les quatre scores de performances rencontrés au cours de la carrière scolaire de l'élève, à savoir : les scores en français et en mathématiques aux évaluations nationales de CE2 et de sixième pour les élèves « à l'heure » et pour ceux ayant un an de retard.
- 30 On constate d'une part que la corrélation liant l'évaluation globale par les enseignants et le score global de performances en début de CP est particulièrement élevée ($r = .58$). D'autre part, le score de performances scolaires en début de CP est assez fortement corrélé aux performances de français et de mathématiques de CE2 et de sixième. De même, les corrélations entre le score global à la grille d'observation en CP et les performances scolaires ultérieures atteignent un niveau satisfaisant et stable tout au long de la scolarité. Toutefois, ces dernières concèdent en moyenne entre 0,12 et 0,15 points comparativement à celles obtenues entre l'évaluation des performances de début de CP et les scores de CE2 et de sixième.

Tableau 5/Table 5

	Score à la grille en début de CP (CS)	Score de performances en début de CP (PS)
Score global de performances en début de CP (PS)	.58	1
Score en français en CE2	.44	.56
Score en mathématiques en CE2	.47	.59
Score en français en 6 ^e	.44	.57
Score en mathématiques en 6 ^e	.43	.58
<i>Note.</i> Toutes les corrélations sont significatives à $p < .0001$.		

Corrélations (rho de Spearman) entre le score global à la grille d'observation renseignée par les enseignants en CP (CS), le score global de performances scolaires en CP (PS) et les scores de performances aux évaluations nationales de français et de mathématiques en CE2 et en 6^e (N = 549)

Correlations (Spearman's rho) between the teacher rating (total score) in 1st Grade (CS), the school performances (total score) in 1st Grade (PS) and the performances at the national assessments in French and Mathematics in 3rd and 6th Grade (N = 5549)

- 31 En résumé, ces premiers résultats indiquent que les performances scolaires de CE2 et de sixième sont davantage liées aux performances scolaires initiales des élèves qu'à

l'évaluation de leurs comportements réalisée par les enseignants. Toutefois, le lien entre l'évaluation des enseignants et les performances scolaires est loin d'être négligeable. De fait, il est intéressant de noter qu'un score composé de 11 items présente toujours, 2 et 5 ans plus tard, une valeur informative pour le moins satisfaisante et assez proche de celle d'un score global composé d'un nombre d'items beaucoup plus grand. Cependant, il convient de s'interroger sur la contribution propre de ces deux évaluations à l'explication des performances et des parcours scolaires des élèves, notamment lorsque les variables sociodémographiques sont prises en compte.

Comparaison du pouvoir prédictif des deux scores à l'entrée au CP (grille d'observation et performances scolaires) sur les performances en CE2 et en sixième

- 32 Huit analyses de régression (modèles linéaires généralisés) ont été conduites afin d'estimer l'importance relative des deux variables explicatives de début CP (CS et PS) sur une variable dépendante (respectivement les quatre scores de performances scolaires en CE2 et en sixième). Pour chaque variable dépendante, deux modèles ont été testés. Le premier modèle (M1) permet d'estimer la part explicative des deux scores du début CP (CS et PS). Le second (M2) ajoute aux deux variables de départ cinq variables sociodémographiques, et permet ainsi d'apprécier dans quelle mesure l'effet de CS persiste lorsqu'on contrôle cet effet par le score de performances initiales (PS) et les variables sociodémographiques.
- 33 Avant d'examiner individuellement les contributions de nos variables à la prédiction des performances de CE2 et de sixième, les huit R2 calculés ont été testés. Ils sont tous significatifs avec un risque inférieur à .0001. Pour chacun des modèles, l'ensemble des variables prises collectivement prédit donc les performances à des niveaux supérieurs au hasard.
- 34 En CE2 (voir tableau 6), on constate (modèle 1) que la part explicative du score de performances en CP est supérieure à celle de l'évaluation comportementale pour les performances en français et en mathématiques. Ces résultats signifient également que CS conserve son pouvoir explicatif indépendamment de PS. Si l'on contrôle le score global en début de CP et les variables sociodémographiques (modèle 2), l'effet propre CS reste stable et significatif en français et en mathématiques. Dans ce modèle, le poids du score de performances de début de CP reste également stable et supérieur à celui de l'évaluation comportementale. Enfin, la part de variance expliquée par le modèle 2 est significativement supérieure à celle du modèle 1⁵ pour les deux domaines de compétences considérés.

Tableau 6/Table 6

	β Modèle 1	β Modèle 2
	Explication du score en français en CE2	
Score de performances en CP (PS)	.46*	.41*

Score comportemental en CP (CS)	.17*	.15*
Explication du score en mathématiques en CE2		
Score de performances en CP (PS)	.49*	.46*
Score comportemental en CP (CS)	.18*	.18*
<i>Note.</i> Pour l'explication du score en français, les pourcentages globaux de variance expliquée sont de 33,1 % ($p < .0001$) pour M1 et de 36,9 % ($p < .0001$) pour M2. Pour l'explication du score en mathématiques, $R^2 = 37,4$ % ($p < .0001$) pour M1 et $R^2 = 39$ % ($p < .0001$) pour M2. * $p < .0001$.		

Poids des variables sociodémographiques, du score de performances scolaires en CP (PS) et du score global à la grille d'observation (CS) sur les performances des élèves aux évaluations nationales de français et de mathématiques en CE2

Weight of the socio-demographic variables, the school performances (total score) in 1st Grade (PS) and the teacher rating (total score) in 1st Grade (CS) on the pupils' performances at the national assessments in French and Mathematics in 3rd Grade

- 35 Ainsi, l'introduction des variables sociodémographiques ajoutées à nos deux prédicteurs, pris collectivement (M2), prédisent significativement mieux le score en français en CE2 que ne le font uniquement nos deux prédicteurs (M1). Malgré l'introduction des variables sociodémographiques, le score comportemental présente un coefficient de régression toujours significatif et apporte une contribution spécifique à l'explication des performances scolaires en français et en mathématiques en CE2.
- 36 En sixième (voir tableau 7), on observe, comme précédemment, que le score de performances initiales contribue davantage que l'évaluation comportementale à l'explication des performances scolaires à 5 ans, qu'il s'agisse du français ou des mathématiques. Toutefois, la part explicative de l'évaluation réalisée par les enseignants reste significative, indépendamment du poids explicatif du score de performances scolaires initial.

Tableau 7/Table 7

	β Modèle 1	β Modèle 2
Explication du score en français en 6 ^e		
Score de performances en CP (PS)	.49*	.44*
Score comportemental en CP (CS)	.17*	.15*
Explication du score en mathématiques en 6 ^e		
Score de performances en CP (PS)	.51*	.46*
Score comportemental en CP (CS)	.13*	.18*

Note. Pour l'explication du score en français, les pourcentages globaux de variance expliquée sont de 35,7 % ($p < .0001$) pour M1 et de 42,2 % ($p < .0001$) pour M2. Pour l'explication du score en mathématiques, $R^2 = 35,4$ % ($p < .0001$) pour M1 et $R^2 = 40,6$ % ($p < .0001$) pour M2. * $p < .0001$.

Poids des variables sociodémographiques, du score de performances scolaires en CP (PS) et du score global à la grille d'observation (CS) sur les performances des élèves aux évaluations nationales de français et de mathématiques à l'entrée en 6^e

Weight of the socio-demographic variables, the school performances (total score) in 1st Grade (PS) and the teacher rating (total score) in 1st Grade (CS) on the pupils' performances at the national assessments in French and Mathematics in 6th Grade

- 37 Bien que le modèle 2 améliore significativement le pourcentage global de variance expliquée (écart de 6,5 % en français, $p < .0001$ et de 5,2 %, $p < .0001$ en mathématiques par rapport au modèle 1), les résultats vont également dans le sens de ceux obtenus en CE2 : stabilité de l'effet propre de CS pour les deux domaines de compétences ainsi que de PS. En définitive, l'évaluation réalisée par les enseignants apporte un supplément d'information dans l'explication de la performance future des élèves, indépendamment des variables sociodémographiques invoquées et de l'évaluation des performances scolaires du début CP. Cette évaluation ne constitue pas la variable la plus explicative des performances ultérieures des élèves, contrairement au score de performances en début de CP ; elle pourrait cependant avoir un pouvoir prédictif supérieur concernant la carrière scolaire des élèves.

Comparaison du pouvoir prédictif des deux scores à l'entrée au CP (grille d'observation et compétences cognitives) sur les parcours scolaires entre le CP et la sixième

- 38 L'étude de la prédiction du parcours scolaire des élèves a été réalisée au moyen de régressions logistiques. L'objectif est de comparer la capacité des deux évaluations à prédire le fait, pour un élève, d'être « à l'heure » ou en retard à l'entrée en CE2 puis à l'entrée en sixième. Il s'agit également de déterminer dans quelle mesure l'effet de l'évaluation réalisée à partir de la grille d'observation persiste lorsque cet effet est contrôlé par le score de performances initiales, puis par les variables sociodémographiques. Comme dans les analyses précédentes, deux modèles ont été testés : le premier (modèle 1) intégrant uniquement les deux scores de début CP, le second (modèle 2) testant simultanément les effets de ces deux facteurs et ceux liés aux variables sociodémographiques.
- 39 Avant d'examiner individuellement les contributions de nos variables à la prédiction des parcours scolaires entre le CP et la sixième, les quatre indicateurs d'ajustement calculés ($-2\log L$), ont été testés. Ils sont tous significatifs avec un risque inférieur à .0001. Pour chacun des modèles, l'introduction de nos variables apporte une contribution significative à la prédiction des parcours scolaires.
- 40 Les données (voir tableau 8) montrent que les coefficients de régression (β) atteignent des valeurs similaires, quels que soient la période prise en compte (CE2 versus sixième) ou le modèle considéré (M1 versus M2). De ce point de vue, l'introduction des facteurs sociodémographiques semble ne pas avoir d'impact sur les coefficients de régression de nos deux prédicteurs (PS et CS) même si les degrés d'amélioration des modèles sont significatifs⁶.

Tableau 8/Table 8

	β Modèle 1	β Modèle 2
	Probabilité d'accès au CE2 avec redoublement	
Score de performances en CP (PS)	- 1,26*	- 1,2*
Score comportemental en CP (CS)	- ,92*	- ,90*
	Probabilité d'accès en 6 ^e avec redoublement	
Score de performances en CP (PS)	- 1,13*	.46*
Score comportemental en CP (CS)	- ,78*	- ,78*
<p>Note. Pour la probabilité d'accès en CE2 avec redoublement, l'ajustement global du modèle (-2LogL) est $\chi^2(2) = 658,5$ ($p < .0001$) pour M1 et $\chi^2(2) = 701,8$ ($p < .0001$) pour M2. Pour la probabilité d'accès en 6^e avec redoublement, $\chi^2(2) = 984,3$ ($p < .0001$) pour M1 et $\chi^2(12) = 1\,040,2$ ($p < .0001$) pour M2.</p> <p>* $p < .0001$.</p>		

Poids des variables sociodémographiques, du score de performances en CP (PS) et du score global à la grille d'observation (CS) sur la probabilité d'accès en CE2 ou en 6^e avec redoublement (régressions logistiques)

Weight of the socio-demographic variables, the school performances score in 1st Grade (PS) and the teacher rating in 1st Grade (CS) on the probability of going to 3rd and 6th Grade after repeating (logistic regression)

- 41 Par ailleurs, le score global de performances scolaires en début de CP apporte, comparativement à l'évaluation comportementale, la contribution la plus forte à l'explication de la trajectoire scolaire des élèves. Toutefois, l'effet propre de l'évaluation réalisée par les enseignants reste significatif aux deux moments de mesure, que les variables sociodémographiques soient prises en compte ou non. Ces résultats vont donc dans le sens de ceux relevés précédemment : le score global à la grille d'observation (CS) contribue à l'explication des performances et des trajectoires scolaires de manière spécifique et indépendante de la part explicative apportée par le score de performances initiales des élèves (PS) et leurs caractéristiques sociodémographiques.

Comparaison de la valeur prédictive de la grille d'observation sur les performances et les parcours scolaires en CE2 et en sixième

- 42 Le tableau 9 regroupe, pour chaque objectif visé (prédiction des trajectoires versus prédiction des performances en CE2 et en sixième) et chaque modèle testé (modèle 1 versus modèle 2), les valeurs des coefficients de régression expliquant la contribution du score de performances en CP (β_{PS}) et celui de la grille (β_{CS}), ainsi que la valeur du ratio calculé de ces deux coefficients (β_{PS} / β_{CS}). Un ratio élevé signifie que le pouvoir explicatif de l'évaluation comportementale est faible comparé à celui obtenu à partir de l'évaluation des performances. Inversement, un ratio faible indique que la prédiction réalisée à partir

du score à la grille d'observation se rapproche de celle issue du score de performances. Ces ratios sont notés de R1 à R12 dans la dernière colonne du tableau 9.

Tableau 9/Table 9

Objectifs	Modèle	β_{PS}	β_{CS}	β_{PS}/β_{CS} (ratio)	Nom du ratio
Explication du redoublement en CE2	M1	- 1,26	- .92	1,4	R1
	M2	- 1,20	- .90	1,3	R2
Explication du redoublement en 6 ^e	M1	- 1,13	- .80	1,4	R3
	M2	- 1,09	- .78	1,4	R4
Explication du score en mathématiques en CE2	M1	.48	.18	2,7	R5
	M2	.45	.17	2,6	R6
Explication du score en français en CE2	M1	.46	.17	2,7	R7
	M2	.41	.15	2,7	R8
Explication du score en mathématiques en 6 ^e	M1	.51	.14	3,7	R9
	M2	.47	.14	3,3	R10
Explication du score en français en 6 ^e	M1	.49	.17	2,9	R11
	M2	.44	.15	2,8	R12

Valeurs des coefficients de régression obtenus pour expliquer la valeur prédictive du score de performances (β_{PS}) et de l'évaluation par les enseignants (β_{CS}) en CP en fonction du niveau scolaire (CE2 *versus* sixième), de l'objectif visé (parcours scolaire *versus* performances scolaires) et du modèle testé (M1 *versus* M2)

*Regression coefficients values obtained to explain the predictive value of the school performances in 1st Grade (β_{PS}) and the teacher rating (β_{CS}) in 1st Grade according to the school level (3rd *versus* 6th grade), the aim of the analysis (school trajectory *versus* performances) and the model tested (M1 *versus* M2)*

- 43 On constate que les ratios correspondant à l'explication des parcours scolaires en CE2 et en sixième (R1 à R4) atteignent tous une valeur proche de 1,4 alors que ceux calculés pour la prédiction des performances scolaires (R5 à R12) varient selon la modalité testée entre 2,6 et 3,7. Comparativement au score de performance en début de CP, la contribution de l'évaluation des enseignants semble donc plus importante dans l'explication des parcours scolaires que dans l'explication de la performance scolaire. Ainsi, à niveau scolaire fixe, on constate que les ratios relatifs au parcours scolaire sont toujours nettement supérieurs à ceux relatifs à la performance scolaire (par exemple, R5 > R1 ou R9 > R3). Est-ce à dire que l'évaluation initiale des enseignants prédit toujours mieux le parcours scolaire des élèves que leurs performances scolaires ?

- 44 Pour répondre à cette question, nous avons, dans un premier temps, soustrait les ratios relatifs à l'explication de la performance (R5 à R12) à ceux relatifs à l'explication du parcours (R1 à R4), ceci à niveaux scolaires, nature et nombre de variables identiques⁷. Huit nouvelles variables ont donc été déterminées (voir tableau 10, première colonne de gauche). Dans un second temps, nous avons interrogé la significativité de ces différences deux à deux. La méthode du *bootstrap* a été utilisée afin d'extraire au hasard 1 000 sous-échantillons de taille identique à l'échantillon initial (N = 5549). Pour chaque nouvelle variable, la moyenne des différences de ratio deux à deux a été calculée sur l'ensemble des 1 000 tirages. Les distributions des moyennes étant toutes d'allures normales⁸, nous avons utilisé la méthode de l'erreur standard pour définir l'intervalle de confiance autour de chacune de ces moyennes des différences de ratio (Palm, 2002). Cette méthode permet en définitive de savoir si le fait que l'évaluation initiale des enseignants prédise mieux le parcours scolaire des élèves que leurs performances est un phénomène anecdotique, propre à notre échantillon, ou à l'inverse, si les chances d'observer ce phénomène sur d'autres échantillons sont suffisamment élevées pour que l'on puisse généraliser ce résultat.

Tableau 10/Table 10

Modalités	Mini	Maxi	Moyenne	Écart-type	L+	L-
CE2 maths M1 (R5-R1)	.27	2,72	1,37	.33	2,02	.72
CE2 français M1 (R7-R1)	.15	3,14	1,37	.35	2,06	.67
CE2 maths M2 (R6-R2)	.24	2,66	1,32	.33	1,98	.66
CE2 français M2 (R8-R2)	.09	3,21	1,39	.37	2,13	.65
6 ^e maths M1 (R9-R3)	1,27	4,12	2,50	.48	3,44	1,56
6 ^e français M1 (R11-R3)	.54	3,45	1,54	.32	2,16	.91
6 ^e maths M2 (R10-R4)	.99	3,58	2,11	.41	2,92	1,30
6 ^e français M2 (R12-R4)	.50	3,02	1,50	.32	2,14	.86

Comparaison de la valeur prédictive de la grille d'observation sur les performances et les parcours scolaires en CE2 et en 6^e au moyen de la méthode du bootstrapping (minimum, maximum, moyenne, écart-type et limites supérieures (L+) et inférieures (L-) de confiance (au seuil de 95 %) des distributions des différences de ratio deux-à-deux)

Comparison of the predictive value of the teacher rating scale on the performances and the school pathways in 3rd and 6th Grade with the bootstrapping method (minimum, maximum, mean value, standard deviation, and the superior (L+) and inferior (L-) limits of the distributions for the ratios differences)

- 45 Les données du tableau 10 montrent, d'une part, que quelle que soit la variable considérée, la moyenne des différences calculée sur les 1 000 tirages est positive au risque de 5 %. Par ailleurs, les ratios calculés pour prédire le parcours scolaire des élèves sont significativement inférieurs à ceux correspondant à la prédiction de leurs performances scolaires et ceci indépendamment du modèle de référence et du niveau scolaire.

L'évaluation initiale des enseignants contribue donc davantage à l'explication du parcours des élèves qu'à celle de leurs performances académiques en CE2 et en sixième.

Discussion-conclusion

- 46 Cette étude montre tout d'abord qu'en début d'école élémentaire, l'évaluation par les enseignants du comportement des élèves en classe est fortement liée aux performances scolaires évaluées à partir d'épreuves standardisées. Ce résultat va tout à fait dans le sens des travaux soulignant la convergence globale de ces deux types d'évaluation (Agostin & Bain, 1997 ; Hoge & Coladarci, 1989 ; Kenny & Chekaluk, 1993 ; Meljac *et al.*, 2001 ; Quay & Steele, 1998) et l'importance des variables comportementales en lien avec les apprentissages pour rendre compte de l'adaptation scolaire des élèves (Alexander *et al.* 1993 ; McClelland *et al.*, 2000).
- 47 Cette étude confirme également ce que de nombreux travaux antérieurs ont pu mettre en évidence à différentes périodes de la scolarité : les évaluations réalisées par les enseignants sont en mesure de prédire les performances académiques ultérieures des élèves (Écalle, 1999 ; Guimard & Florin, 2001 ; Mantzicopoulos & Morrison, 1994 ; Quay & Steele, 1998). Notre étude longitudinale souligne en outre que le niveau des prédictions reste stable quels que soient le niveau scolaire étudié et le type de compétences évaluées.
- 48 Conformément aux résultats de l'étude transversale de Yen *et al.* (2004), on constate que les performances académiques des élèves de CE2 et de sixième sont mieux prédites par l'évaluation standardisée initiale que par celle du questionnaire enseignant. Les analyses de régression indiquent toutefois que l'évaluation des enseignants apporte une contribution significative à l'explication des performances scolaires ultérieures indépendamment de l'effet propre de l'évaluation standardisée initiale et des variables sociodémographiques. Notons que cet effet reste significatif même lorsque la prédiction est réalisée à 5 ans, ce qui atteste de la stabilité de la valeur prédictive de l'évaluation par les enseignants sur la période considérée.
- 49 Une des originalités de cette recherche était également de tester la capacité de l'évaluation magistrale à prédire la réussite ou les difficultés d'adaptation scolaire des élèves au cours de leur scolarité élémentaire. Les régressions logistiques indiquent que « toutes choses égales par ailleurs », le parcours ultérieur des élèves est mieux expliqué par l'évaluation initiale des performances scolaires que par celle des enseignants. Cependant, là encore, l'évaluation réalisée par les enseignants reste déterminante aux deux périodes étudiées, même lorsque les performances scolaires initiales et les caractéristiques sociodémographiques sont contrôlées. De plus, les évaluations comportementales semblent significativement mieux expliquer le parcours des élèves que leurs performances scolaires.
- 50 Toutefois, la généralisation de ces dernières observations doit tenir compte du fait que l'échantillon étudié ne présente pas l'ensemble des caractéristiques sociodémographiques de la population globale. Il est à noter, en particulier, que les élèves les plus susceptibles d'être en difficulté scolaire (élèves étrangers et appartenant à des catégories socioprofessionnelles défavorisées) sont sous-représentés dans notre échantillon, ce qui conduit mécaniquement à une plus faible proportion de redoublants.
- 51 En résumé, ces résultats indiquent clairement que l'évaluation par les enseignants des comportements et des compétences des élèves en classe apporte un gain d'information,

par rapport aux évaluations des performances scolaires standardisées, dans la prédiction des performances scolaires et des trajectoires de réussite ou d'échec des élèves. De ce fait, parce qu'elle apporte des informations complémentaires à celles provenant des évaluations cognitives standardisées, une évaluation de ce type pourrait s'insérer dans les protocoles d'évaluation plus larges destinés à mieux comprendre le fonctionnement des élèves en classe et à mieux anticiper leurs éventuelles difficultés d'adaptation à l'école (Flynn & Rahbar, 1998 ; Mantzicopoulos & Morrison, 1994 ; Yen *et al.*, 2004).

- 52 Ceci ne signifie en aucun cas qu'une telle évaluation doive se substituer aux évaluations cognitives. En effet, comme on l'a montré dans cette étude et comme l'ont mis en évidence diverses recherches (Belsito, Ryan & Brophy, 2005 ; Cadieux & Boudreault, 2002 ; Crosby & French, 2002 ; Guimard & Florin, 2001 ; Teisl *et al.*, 2001 ; Yen *et al.*, 2004), ces dernières contribuent fortement à la prédiction des compétences et des parcours scolaires. Par ailleurs, les évaluations que réalisent les enseignants ne permettent ni de déterminer les processus en jeu dans les apprentissages, ni d'aboutir à un diagnostic pour les élèves en difficulté. Pour autant, elles sollicitent une description de comportements et de savoir-faire des enfants dans les tâches et les activités scolaires tels que l'enseignant peut les repérer au quotidien. Elles peuvent donc aider l'enseignant à prendre en compte certaines dimensions comportementales auxquelles il n'aurait pas été spontanément sensible, au moins pour certains enfants de sa classe. Elles peuvent ainsi constituer un moyen simple et utile de repérage des difficultés des enfants en classe, en vue d'établir un dialogue avec les parents, les autres enseignants et les professionnels concernés (psychologues, médecins scolaires, par exemple).
- 53 La connaissance que les enseignants ont du fonctionnement de leurs élèves en classe, de leurs savoir-faire et de leurs difficultés peut contribuer à expliquer les résultats obtenus (O'Neill & Liljequist, 2002 ; Zazzo, 1978). C'est pourquoi des évaluations de ce type sont utilisées régulièrement dans d'autres pays, notamment anglo-saxons. Souhaitons que certaines résistances à l'évaluation puissent être dépassées et qu'il en soit de même en France, dès l'école maternelle, pour une meilleure prévention et un meilleur accompagnement des difficultés scolaires des enfants. Car à défaut de pouvoir tout prédire (ce qui serait un déterminisme absolu démenti par la psychologie du développement), il est possible de prévenir les difficultés et de mettre en place des aides au développement des enfants, lorsque c'est nécessaire. Selon les difficultés renseignées par les enseignants dans ce type de questionnaire et avec des évaluations standardisées complémentaires si nécessaire, des aides spécifiques peuvent être mises en place précocement, dans la classe ou en dehors.

BIBLIOGRAPHIE

Agostin, T. M. & Bain, S. K. (1997). Predicting early school success with developmental and social skills screeners. *Psychology in the Schools*, 34 (3), 219-228.

Alexander, K. R., Entwisle, D. R. & Dauber, S. L. (1993). First grade classroom behavior: its short and long-term consequences for school performance. *Child développement*, 64, 801-814.

- Belsito, L., Ryan, B. A. & Brophy, K. (2005). Using behavioral and academic indicators in the classroom to screen for at-risk status. *Psychology in the Schools*, 42 (2), 151-158.
- Cadieux, A. & Boudreault, P. (2002). Psychometric properties of a Kindergarten Behavior Rating Scale to predict later academic achievement. *Psychological Reports*, 90 (2), 687-698.
- Caille, J.-P. (2001). Scolarisation à 2 ans et réussite de la carrière scolaire au début de l'école élémentaire. *Éducation et formation*, 60, 3-15.
- Caille, J.-P. (2004). Le redoublement à l'école élémentaire et dans l'enseignement secondaire : évolutions des redoublements et parcours scolaires des redoublants au cours des années 1990-2000. *Éducation et formation*, 69, 79-88.
- Chevrie-Müller, C. (1999). Questionnaire « Langage et comportement – 3 ans ½ ». *Les cahiers pratiques de l'ANAE*, 1-16.
- Coleman, J. M. & Dover G. M. (1993). The RISK screening test: using Kindergarten teacher' ratings to predict future placement in resource classrooms. *Exceptional Children*, 59 (5), 468-477.
- Cosnefroy, O. & Rocher, T. (2004). Le redoublement au cours de la scolarité obligatoire : nouvelles analyses, mêmes constats. *Éducation et formation*, 70, 73-82.
- Crahay, M. (2003). *Peut-on lutter contre l'échec scolaire ?* Bruxelles : De Boeck Université.
- Crahay, M. (2004). Peut-on conclure à propos des effets du redoublement ? *Revue française de pédagogie*, 148, 11-23.
- Crosby, E. G. & French, J. L. (2002). Psychometric data for teacher judgments regarding the learning behaviors of primary grade children. *Psychology in the Schools*, 39 (3), 235-244.
- DiLalla, L. F., Marcus, J. L. & Wright-Phillips, M. V. (2004). Longitudinal effects of preschool behavioral styles on early adolescent school performance. *Journal of School Psychology*, 42, 385-401.
- Duru-Bellat, M. (2003). Actualité et nouveaux développements de la question de la reproduction des inégalités sociales. *L'Orientation scolaire et professionnelle*, 4, 571-594.
- Écalle, J. (1999). Prédiction de la réussite scolaire en lecture-écriture au cycle II. *Revue européenne de psychologie appliquée*, 50 (1), 81-85.
- Ferrier, J. (2003). *L'avance et le retard scolaires à l'école élémentaire et au collège*. Document non publié.
- Fletcher, J. M. & Satz, P. (1984). Test-based versus teacher-based predictions of academic achievement: a three year longitudinal follow-up. *Journal of Pediatric Psychology*, 9 (2), 193-203.
- Fletcher, J., Tannock, R. & Bishop, D. V. M. (2001). Utility of brief teacher rating scales to identify children with educational problems: experience with an Australian sample. *Australian Journal of Psychology*, 53 (2), 63-71.
- Florin, A. (1991). *Pratiques du langage à l'école maternelle et prédiction de la réussite scolaire*. Paris : Presses universitaires de France.
- Florin, A., Guimard, P. & Nocus, I. (2002). Les évaluations des enseignants et la prédiction des compétences langagières de leurs élèves : études longitudinales à l'école maternelle et à l'école élémentaire. *Le langage et l'homme*, 2, 175-190.
- Florin, A., Cosnefroy, O. & Guimard, P. (2004). Trimestre de naissance et parcours scolaires. *Revue européenne de psychologie appliquée*, 54, 237-246.
- Flynn, J. M. & Rahbar, M. H. (1998). Improving teacher prediction of children at risk for reading failure. *Psychology in the schools*, 35 (2), 163-172.

- Foulks, B. & Morrow, R. D. (1989). Academic survival skills for the young child at risk for school failure. *Journal of Educational Research*, 82, 158-165.
- Goldstein, H. & Rabash, J. (2003). *Analysis of 1997 French pupil panel study data* (rapport non publié pour le ministère de l'Éducation nationale).
- Guimard, P. & Florin, A. (2001). Comportements scolaires en moyenne section de maternelle et prédiction de la réussite scolaire à l'école élémentaire. *Psychologie et psychométrie*, 22 (1), 75-100.
- Guimard, P., Florin, A. & Nocus, I. (2002). Comment les enseignants d'école maternelle peuvent-ils prédire les trajectoires scolaires de leurs élèves ? *Revue européenne de psychologie appliquée*, 52 (1), 63-76.
- Hecht, S. A. & Greenfield, D. B. (2002). Explaining the predictive accuracy of teacher judgments of their students' reading achievement: the role of gender, classroom behavior and emergent literacy skills in a longitudinal sample of children exposed to poverty. *Reading and Writing: An Interdisciplinary Journal*, 15, 789-809.
- Hoge, R. D. (1983). Psychométric Properties of Teacher-Judgment Measures of Pupil Attitudes, Classroom Behaviors and Achievement Levels. *The journal of Special Education*, 17 (4), 401-429.
- Hoge, R. D. & Coladarci, T. (1989). Teacher-based judgments of academic achievement: a review of the literature. *Review of Educational Research*, 59, 297-313.
- Howell, D. C. (1998). *Méthodes statistiques en sciences humaines*. Bruxelles : De Boeck Université.
- Johnson, R. W. (2001). An introduction to the bootstrap. *Teaching Statistics*, 23, 49-54.
- Jeantheau, J.-P. & Murat, F. (1998). Observation à l'entrée au CP du « panel 1997 » (note d'information n° 98-40). Paris : ministère de l'Éducation nationale, de la Recherche et de la Technologie.
- Kenny, D. T. & Chekaluk, E. (1993). Early reading performance: a comparison of teacher-based and test-based assessments. *Journal of learning disabilities*, 26 (4), 227-236.
- Ladd, G. W., Birsh, S. H. & Buhs, E. S. (1999). Children's social and scholastic lives in kindergarden: related spheres of influence? *Child development*, 70, 1373-1400.
- Laveault, D. & Grégoire, J. (1997). *Introduction aux théories des tests en sciences humaines*. Bruxelles : De Boeck Université.
- Le Bastard, S. & Suchaut, B. (2000). Lecture-écriture au cycle II. Évaluation d'une démarche innovante. *Les cahiers de l'IREDU*, 61.
- McClelland, M. M., Morisson, F. J. & Holmes, D. L. (2000). Children at risk for early academic problems: the role of learning-related social skills. *Early Childhood Research Quarterly*, 15, 307-329.
- McClelland, M. M. & Morisson, F. J. (2003). The emergence of learning-related social skills in preschool children. *Early Childhood Research Quarterly*, 18, 206-224.
- McCloskey, G. (1990). Selecting and Using Early Childhood Rating Scales. *Topics in Early Childhood Special Education*, 10 (3), 39-64.
- McDermott, P. A. (1999). National scales of differential learning behaviors among American children and adolescents. *School Psychology Review*, 28, 280-291.
- McDermott, P. A., Leigh, N. M. & Perry, M. A. (2002). Development and validation of the preschool learning behavior scales. *Psychology in the Schools*, 39, 353-365.

- Mantzicopoulos, P. Y. & Morrison, D. (1994). Early prediction of reading achievement. Exploring the relationship of cognitive and noncognitive measures to inaccurate classifications of at-risk status. *Remedial and special education*, 4, 244-251.
- Meljac, C., Kugler, M. & Mogenet, J.-L. (2001). Prévion, prédiction, anticipation et apprentissage de la lecture. *Pratiques psychologiques*, 1, 81-97.
- Meisels, S. J., Bickel, D. D., Nicholson, J., Xue, Y. & Atkins-Burnett, S. (2001). Trusting teachers' Judgments: a validity study of a curriculum-embedded performance assessment in Kindergarten to Grade 3. *American Educational Research Journal*, 38 (1), 73-95.
- Mercer, C. D., Algozzine, B. & Trifiletti, J. (1988). Early identification: an analysis of the research. *Learning Disability Quarterly*, 11, 176-188.
- Mingat, A. (1984). Les acquisitions scolaires au CP : les origines des différences. *Revue française de pédagogie*, 69, 49-62.
- Mingat, A. & Richard, M. (1991). Évaluation des activités de rééducation GAPP à l'école primaire. *Les cahiers de l'IREDU*, 49.
- Ministère de l'Éducation nationale (MEN) (2000). Évaluations à l'entrée au CE2 et en sixième. *Repères nationaux*, n° hors série.
- Ministère de l'Éducation nationale (MEN) (2002). *Les compétences des élèves à l'entrée au cours préparatoire. Études réalisées à partir du panel d'écoliers recruté en 1997*. Direction de la programmation et du développement, n° 132.
- Ministère de l'Éducation nationale (MEN) (2003). Évaluations à l'entrée au CE2 et en sixième. *Repères nationaux*, n° hors série.
- Ministère de l'Éducation nationale (MEN) (2005). *Repères et références statistiques sur les enseignements, la formation et la recherche*.
- Oehler-Stinette, J. & Boykin, C. (2001). Convergent, Discriminant and predictive validity of the teacher rating of academic achievement motivation (TRAAM) with the ACTeRS-TF and the BASC-TRS. *Journal of the Psychoeducational Assessment*, 19, 4-18.
- O'Neill, K. B. & Liljequist, L. (2002). Strategies used by teachers to rate student behavior. *Psychology in the schools*, 39 (1), 77-85.
- Palm, R. (2002). Utilisation du bootstrap pour les problèmes statistiques liés à l'estimation des paramètres. *Biotechnologie, agronomie, société et environnement*, 6, 143-153.
- Quay, L. C. & Steele, D. C. (1998). Predictive children achievement from teacher judgements: an alternative to standardized testing, *Early Education and Development*, 9 (3), 207-218.
- SAS Institute Inc. (1990). *SAS Language: Reference*, version 6, 1^{re} éd.
- Taylor, H. G., Anselmo, A., Foreman, A. L., Schatschneider, C. & Angelopoulos, J. (2000). Utility of kindergarten teacher judgments in identifying early learning problems. *Journal of Learning Disabilities*, 33 (2) 200-210.
- Teisl, J. T., Mazzocco, M. M. & Myers, G. F. (2001). The utility of kindergarten teacher ratings for predicting low academic achievement in first grade. *Journal of Learning Disabilities*, 34 (3), 286-293.
- Yen, C. J., Konold, T. R. & McDermott, P. A. (2004). Does learning behavior augment cognitive ability as an indicator of academic achievement? *Journal of School Psychology*, 42, 157-169.
- Zazzo, B. (1978). *Un grand passage : de l'école maternelle à l'école élémentaire*. Paris : Presses universitaires de France.

NOTES

1. Les catégories socioprofessionnelles sont constituées à partir de la profession des parents en ne retenant que le meilleur classement. La catégorie « favorisée » regroupe les enfants d'enseignants (professeurs ou instituteurs), de cadres supérieurs et de personnes exerçant une profession libérale. La catégorie « plutôt favorisée » correspond aux professions intermédiaires. La catégorie « moyenne » est celle des employés, des artisans-commerçants et des agriculteurs. Enfin, les ouvriers, les inactifs constituent la catégorie « défavorisée ».

2. Direction de l'évaluation de la prospective et des performances.

3. Ce questionnaire comportait initialement 15 questions. Les items 14 et 15 n'ont pas été retenus car, contrairement aux treize autres, ils présentent des échelles de mesure différentes.

4. Pour comparer l'homogénéité de plusieurs échelles de taille différente, on calcule α_{st} avec k correspondant à la proportion de la modification du nombre d'items. En notant \bar{r} la fidélité du test initial, on a :

$$\alpha_{st} = \frac{k\bar{r}}{1 + (k-1)\bar{r}}$$

5. La statistique partielle de Fisher a été utilisée afin de comparer les deux modèles emboîtés. Nous interrogeons la significativité de la différence de $R_{M2} - R_{M1}$ lorsque l'on rajoute les 5 variables sociodémographiques, soit, avec p correspondant aux 7 variables du modèle M2 :

$$R_{M2} - R_{M1} > \frac{q(1 - R_{M1})}{(n - p - 1)} F_{q,q;(n-p-1)}$$

Ainsi, par exemple, pour comparer les deux modèles emboîtés en français, on obtient, avec $F_{.05(5,5541)} = 2,21$; $p < 0.0001$:

$$\frac{(R_{M2} - R_{M1})}{(1 - R_{M1})} \frac{n - p - 1}{q} = \frac{(0,37 - 0,33)}{(1 - 0,37)} \frac{5\,549 - 7 - 1}{5} = 66,5$$

6. Le degré d'amélioration entre les deux modèles a été déterminé en calculant la différence des ajustements des deux modèles (Howell, 1998). On constate que χ^2_{M2-M1} avec (12-2) degrés de liberté vaut respectivement 43,3 pour la modélisation de l'accès en CE2 avec redoublement et 55,9 pour la modélisation de l'accès en sixième avec redoublement. La valeur critique du chi-carré au seuil de 5 % avec 10 ddl est de 18,3, $p < .0001$.

7. Par exemple, au niveau CE2, lorsque l'on s'intéresse aux modèles ne comportant que les deux scores (CS et PS) nous pouvons comparer R1 à R5 et R7.

8. Test de normalité de Kolmogorov-Smirnov effectué sur les 8 distributions.

RÉSUMÉS

La présente étude vise à comparer la capacité d'une évaluation des comportements et des compétences des élèves effectuée par les enseignants et d'une évaluation cognitive standardisée à prédire les performances et les parcours scolaires ultérieurs. 5 549 élèves provenant d'un panel

constitué par la Direction de l'évaluation et de la prospective du ministère de l'Éducation nationale (MEN) ont été suivis longitudinalement de la première année primaire (cours préparatoire) à la sixième de collège. En début de première année primaire, les comportements en classe et les compétences des élèves ont été évalués par leurs enseignants, et leurs performances scolaires au moyen d'un protocole standardisé. Ces données ont été mises en relation avec les résultats de ces mêmes élèves aux évaluations nationales de CE2 et de sixième en français et en mathématiques ainsi qu'avec les données concernant leur parcours scolaire (redoublement ou non). Les analyses (régressions linéaires et logistiques) montrent que l'évaluation des enseignants apporte une contribution significative à l'explication des performances et des trajectoires scolaires ultérieures indépendamment de l'effet propre des performances scolaires initiales et des variables sociodémographiques. Les implications de cette recherche dans le champ de l'évaluation en psychologie de l'éducation et de la prévention des difficultés d'apprentissage sont présentées.

The current study compares the capacity of a teacher rating and of a cognitive standardized evaluation to predict the future pupils' academic performances and their later school pathways. 5 549 pupils coming from a panel built up by the Evaluation and Forecasting Directorate from the French Ministry of National Education, were followed longitudinally from 1st Grade to 6th Grade. At the beginning of 1st Grade, the pupils' behaviours were evaluated by their teachers, and their academic performances by means of a standardized protocol. These data were linked with the results of these pupils at the national assessments in 3rd Grade and in 6st Grade in French and Maths and with the data concerning their school pathway (repeating or not). The analyses (linear and logistic regressions) show that teacher rating contributes significantly to the explanation of the future academic performances and later school pathways independently of the specific effect of the initial academic performances and the socio-demographic variables. The implications of this research in the field of the psycho-educational assessment and the prevention of learning difficulties are presented.

INDEX

Mots-clés : Comportements et compétences scolaires, Évaluation par les enseignants, Prédiction de la réussite scolaire, Prévention des difficultés d'apprentissage

Keywords : Prediction of school success, Prevention of learning difficulties, School behaviours and competences, Teacher rating

AUTEURS

PHILIPPE GUIMARD

Philippe Guimardest maître de conférences en psychologie du développement à l'université de Nantes (EA 3259). Thèmes de recherche : liens oral-écrit, influence des contextes éducatifs sur les apprentissages, prévention des difficultés d'apprentissage, dimensions conatives en éducation. Contact : Laboratoire de psychologie « Éducation, cognition, développement » (LabÉCD, EA 3259), université de Nantes, BP 81227, 44036 Nantes Cedex 3. Courriel : philippe.guimard@univ-nantes.fr

OLIVIER COSNEFROY

Olivier Cosnefroy, ancien chargé de mission à la Direction de l'évaluation, de la prospective et des performances (DEPP) du ministère de l'Éducation nationale, est, depuis 2007, ingénieur d'études

au laboratoire de psychologie (EA 4139) de l'université de Bordeaux 2. Thèmes de recherche : inégalités scolaires, développements cognitifs et comportementaux : suivis longitudinaux d'élèves à l'école primaire (thèse en cours).

AGNÈS FLORIN

Agnès Florin est professeur des universités en psychologie du développement et de l'éducation à l'université de Nantes (EA 3259). Thèmes de recherche : interactions adultes-enfants et développement psychologique, évaluation des compétences des jeunes enfants, prévention des difficultés d'apprentissage et de l'illettrisme, dimensions conatives en éducation.