

CONFÉRENCE DE C O N S E N S U S

LIRE, COMPRENDRE, APPRENDRE :
COMMENT SOUTENIR LE DÉVELOPPEMENT
DES COMPÉTENCES EN LECTURE ?

Les compétences des élèves français
en lecture et compréhension

Mars 2016



LES COMPÉTENCES DES ÉLÈVES FRANÇAIS EN LECTURE ET COMPRÉHENSION :

Un complément à l'article de Daussin et *al* (2011)

Cnesco

Synthèse réalisée pour la préparation de la conférence de consensus « Lire, comprendre, apprendre : comment soutenir le développement de compétences en lecture ? »

Cnesco-Ifé, Lyon, 16-17 mars 2016

Introduction

Dans l'organisation de la conférence de consensus sur le thème « *Lire, comprendre, apprendre : comment soutenir le développement de compétences en lecture?* », le Cnesco et l'Ifé-ENS de Lyon ont produit un ensemble de contributions visant à identifier, analyser et synthétiser la littérature traitant des questions d'apprentissage et d'enseignement de la lecture tout au long de la scolarité. Une première étape importante consiste à réaliser un bilan des compétences des élèves en France et de l'évolution de leurs performances. C'est le rôle de cette note.

Dans l'édition 2011 de la publication de l'INSEE portrait social, on trouve l'article de Daussin, Keskaik et Rocher intitulé : L'évolution du nombre d'élèves en difficulté face à l'écrit depuis une dizaine d'années.

Ce travail présente les principaux apports et les principaux résultats des dispositifs d'évaluation des élèves en France dans les domaines de la lecture et de l'écriture. Ils s'appuient pour cela sur les enquêtes menées par la DEPP (Cedre ; Lire, écrire compter ; SPEC6) et les programmes internationaux auxquels la France participe (PIRLS et PISA).

Ces enquêtes ne mesurent pas les performances des élèves au même moment de leur scolarité et les contenus de ces tests les rendent plus ou moins informatifs sur le niveau des élèves, suivant leur rang dans la distribution des performances. Nous précisons ces éléments dans la section suivante, puis nous prolongerons les résultats au-delà de 2011 à partir d'évaluations réalisées ensuite.

Comment mesurer les compétences en lecture ?

Les compétences des élèves sont par essence une notion complexe, multidimensionnelle et inobservable. Seules leurs manifestations peuvent être appréhendées, par exemple à travers les résultats à un test. Cependant, celles en lecture et en compréhension peuvent apparaître particulièrement sensibles à mesurer.

Les programmes d'évaluation mobilisent conjointement deux champs disciplinaires : la psychométrie et la statistique. Le premier vise à mesurer efficacement les acquis cognitifs des élèves, le second à permettre les conditions de collectes de données qui permettent d'obtenir des résultats fiables.

Puisque les compétences ne peuvent être directement observées, l'idée générale de la psychométrie est de construire et d'exploiter un test dont le contenu, à travers la variation de probabilité de succès de chaque item, mesure une **variable latente** qui influence directement les chances de réussite aux items et, qui, en quelque sorte, s'apparente à la compétence testée. Tout se passe alors comme si on supposait que la compétence mesurée est distribuée parmi les élèves selon une loi statistique connue¹ pour laquelle un ensemble de questions bien construites, administrées à un échantillon large et représentatif, permet d'en identifier les principales caractéristiques et propriétés.

Pour que le teste mesure bien la compétence visée, les éléments du test doivent être **valides**, c'est-à-dire mesurer effectivement ce qu'ils sont censés mesurer.

¹ Et même contrainte dans les modèles. Par exemple un modèle de Rasch, tel qu'utilisé dans les tests PISA, impose une distribution logistique.

Or il n'est pas simple d'assurer la validité d'un test lorsqu'on ne dispose pas de mesure de référence. Comme il n'existe pas de mesure objective de la compréhension en lecture, comment s'assurer que les questions mesurent bien la compréhension et non la connaissance du vocabulaire par exemple ?

Plusieurs sources peuvent être mobilisées. Souvent, la validité est pourvue en faisant appel à des experts pour produire des items qui garantissent la correspondance avec les programmes, avec les attentes à un niveau donnée, etc. C'est notamment le cas pour les évaluations Cedre de la Depp.

Ensuite, le test doit être de la bonne dimension. Cette notion de **dimensionnalité** signifie qu'il faut que les items soient suffisamment liés entre eux, qu'ils ne dépendent que d'un unique facteur latent. Si les items dépendent de plusieurs dimensions, il faut qu'un facteur commun soit prépondérant dans la détermination des réponses. Cette question de la **dimensionnalité** est également liée au problème de **fonctionnement différentiel d'items**. Cette difficulté apparaît lorsque deux groupes de compétences égales ne sont pas capables de répondre correctement à un item de la même façon. Dit autrement, un groupe n'a pas une probabilité égale d'avoir une réponse juste alors que ses membres ont le même niveau de compétence que l'autre groupe. Si le facteur qui cause ce décalage ne fait pas partie des dimensions que l'on souhaite faire apparaître, alors le test est biaisé. L'analyse des fonctionnements différentiels d'items est donc cruciale pour s'assurer de la validité d'un test.

Enfin, des modélisations statistiques de la probabilité de succès aux items permettent ensuite d'établir des scores qui, suivant la modélisation choisie, permettent des comparaisons temporelles. Ce type de modélisation s'inscrit dans le champ de la **théorie des réponses à l'item (MRI, ou IRT, Item response theory)**².

La comparabilité dans le temps est permise par une phase appelée **equating**. Elle requiert qu'une partie (au moins) des items soit reprise dans les questionnaires d'années en années de façon à positionner sur la même échelle de compétence les élèves des différentes vagues d'enquête. Il existe de nombreuses méthodes pour y parvenir, mais, de façon simple, elles consistent généralement à contraindre les paramètres de difficulté des items de la seconde vague d'enquête à prendre les valeurs estimées lors de la vague précédente. Plus concrètement, cela revient à faire l'hypothèse que le niveau de difficulté de l'item n'a pas changé entre les deux périodes.

Cette hypothèse peut sembler triviale aux premiers abords mais en réalité, elle peut s'avérer assez héroïque dans certains contextes. A titre d'exemple, imaginons qu'on administre un test mesurant les compétences en code des enfants de 6 ans. On identifie alors les paramètres de difficulté des items et les scores des élèves. Supposons qu'on reproduise cet exercice plus de 10 ans après en reprenant les mêmes items et en leur attribuant la même valeur de difficulté. Les changements de pratiques en maternelle – qui se seraient par exemple plutôt *primarisés* sur la période et auraient favorisé l'apprentissage du code – s'apparenteraient alors à un changement de difficulté des items. Ce problème n'est pas très différent de celui du fonctionnement différentiel d'items. Tout se passerait comme si le groupe de la deuxième vague réussissait mieux les items de code car, à compétence égale, ils seraient davantage entraînés à y répondre. La comparaison dans le temps sur cette base serait alors biaisée.

² Pour une présentation plus détaillée, voir Rocher (2015) : Mesure des compétences. Méthodes psychométriques utilisées dans le cadre des évaluations des élèves. In DEPP, *Education et Formations, Evaluation des acquis : principes, méthodologie, résultats*. N°86-87. Mai 2015, Paris, MENESR-DEPP.

Il est alors clair que le choix des items – en particulier leur niveau de difficulté – est déterminant dans la précision du score finale et qu'alors, un test de complexité moindre sera plus informatif pour les élèves faibles que pour les élèves forts. Ainsi, les tests proposés dans Lire, écrire compter sont très informatifs pour les élèves faibles, mais beaucoup moins précis pour les élèves plus forts. Les évaluations Cedre CM2 et PIRLS en CM1 sont très efficaces pour les élèves médians, mais le sont moins pour les *bons* ou les *mauvais* élèves.

Tableau 1: Évaluations standardisées en lecture en France adaptées à des comparaisons temporelles valides.

Test	Nom	Années	Population	Domaine
Evaluations nationales				
Cedre	Cycle des évaluations disciplinaires Réalisées sur échantillons	Annuel Depuis 2003 Tous les 5 ans dans une discipline donnée	CM2 et 3 ^e	Maitrise de la langue (MDL) en CM2, compétences générales en 3 ^e , langues vivantes, attitudes à l'égard de la vie en société, histoire, géographie et éducation civique, sciences expérimentales, mathématiques
Lolf	Evaluations pour les indicateurs de la LOLF	Annuel depuis 2007	CE1 CM2 et 3 ^e	Compétences de base en français et en mathématiques, compétences du socle commun
LEC	Lire écrire compter	1987, 1997, 2007	CM2	Compréhension de l'écrit orthographe et calcul.
SPEC6	Etude spécifique des difficultés de lecture	1997 ; 2007	Début de 6 ^e	Automatismes, lexique, compréhension
Panel CP	Evaluations standardisées des acquis des élèves du panel CP	1997 ; 2011	Début de CP	Pré-lecture, écriture, numération, compréhension orale
JDC	Journée défense citoyenneté	Annuel depuis 1998	Environ 17ans	Compréhension de l'écrit, lexique, automatismes
IVQ	Information et vie quotidienne	2004 ; 2011	16 à 65 ans	Littératie, numératie
Evaluations internationales				
PIRLS	Progress in international reading literacy study	2001; 2006; 2011; 2016	CM1	Comprehension de l'écrit
PISA	Program for international student assessment	Tous les 3 ans depuis 2000	15 ans	Compréhension de l'écrit, culture mathématique, culture scientifique
PIAC	Program for the international assesment of adulte competencies	2012	16-65 ans	Littératie et numératie

Source : DEPP (2015) p.29

Le tableau 1 présente l'ensemble des enquêtes menées en France qui concerne la lecture. La plupart des résultats de ces études sont analysés dans Daussin et *al.* 2011. Nous complétons leurs résultats en présentant les enquêtes plus récentes.

Les conclusions des enquêtes avant 2009

La stabilité en moyenne...

Entre 1997 et 2007, les résultats moyens des élèves aux évaluations standardisées, peu importe le niveau auquel on teste les élèves (CP-CM1-CM2-6^e 3^e), sont restés globalement stables. Au primaire, les évaluations Cedre en CM2 présentent une grande stabilité pour tous les élèves, quelle que soit leur place dans la distribution. Pour PIRLS, bien que la moyenne des élèves scolarisés en France soit en deçà de la moyenne des pays de l'union européenne ayant participé, elle est stable entre 2001 et 2006. En fin de 3^e, l'évaluation Cedre montre également une grande stabilité des élèves situés aux niveaux médians, l'enquête PISA des élèves de 15 ans fournit des résultats moyens non significativement différents entre 2000 et 2009. Cependant, la distribution des performances a beaucoup évolué et la stabilité en moyenne cache de fortes disparités dans les *queues de distribution*.

...dissimule une plus forte inégalité de compétences

En effet, dans Cedre la part des élèves situés dans les niveaux les plus élevés a diminué de façon significative, passant de 10,0 % à 7,1 % entre 2003 et 2009. Mais c'est auprès des élèves en difficulté que la dégradation est la plus forte, que ce soit au primaire ou au collège.

Ainsi, le pourcentage d'élèves faibles en compréhension de l'écrit en fin d'école a doublé entre 1997 et 2007, passant de 11 % à 21,4 % (LEC). Pour les compétences d'entrée en 6^e, ces statistiques passent de 14,9 % à 19,0 % entre 1997 et 2007. Ces deux enquêtes sont très informatives sur les niveaux des élèves faibles. En revanche, PIRLS et Cedre – qui sont plus adaptés aux élèves médians – ne montrent pas de dégradation, mais leurs tests sont moins à même de détecter correctement les compétences des élèves faibles.

En fin de collège, les conclusions de Cedre et de PISA sont concordantes et indiquent une augmentation du nombre d'élèves en difficulté et une dégradation de leur niveau. La part d'élèves en difficultés passe de 15,0 % à 17,9 % pour la première et de 15,2 % à 17,9 % pour la seconde.

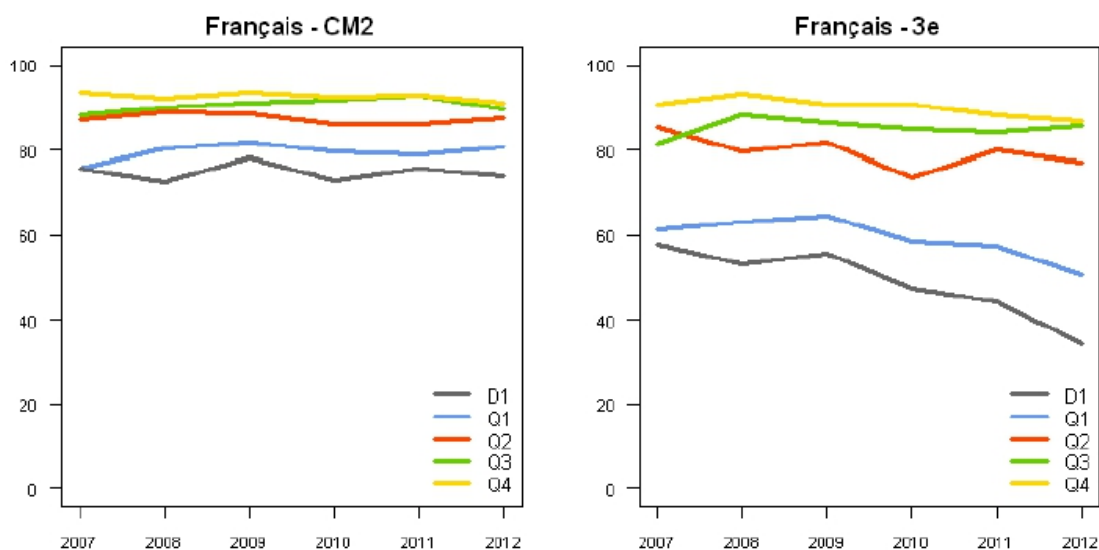
L'origine sociale des élèves et le genre semble aussi impacter plus fortement les compétences en lecture et compréhension des élèves dans les enquêtes récentes.

Les indicateurs de LOLF en maîtrise du Français

Si la maîtrise du Français ne correspond pas exactement à la thématique de la conférence de consensus, elle s'en approche. Aussi, nous ne présentons que brièvement ce résultat à travers un graphique tiré de Rocher (2016)³. Cette analyse compare le pourcentage de compétence acquises telles que définies dans la LOLF en fonction de l'indice de composition sociale des établissements.

³ Rocher (2016, à paraître) : Evolution des inégalités sociales de compétence : une synthèse in Cnesco : *Les inégalités scolaires d'origine sociales et ethnoculturelle : une possible amplification ?* Paris, Cnesco.

Cet indice prend des valeurs faibles lorsque l'établissement est composé d'élèves d'origine sociale majoritairement défavorisée et forte lorsque l'origine sociale des élèves est élevée.



Ces scores sont basés sur des questionnaires identiques reproduits chaque année auprès d'un échantillon représentatif d'élèves. L'indicateur D1 représente le pourcentage de compétence acquise pour les 10 % d'établissements dont la composition sociale est la plus défavorisée. Les scores Q2 à Q4 sont les scores pour les différents quintiles de composition sociale (Q1=20 d'établissements les plus défavorisés, Q2 40 %,...). En fin de CM2, les élèves d'établissements les plus défavorisés maîtrisent environ 75 % des compétences attendues. Cette statistique est stable sur la période. Pour les 20 % d'élèves les plus favorisés, près de 95 % des compétences sont acquises. Ces scores sont stables et à des niveaux relativement satisfaisants, même pour les écoles les plus défavorisées. En revanche, en fin de 3^e, les élèves des 10 % d'établissements les plus défavorisés ne maîtrisaient que 60 % des compétences attendues en 2007 et ce score a chuté à 35 % en 2012 ! En parallèle, les élèves des 60 % d'établissement les plus favorisés (i.e. Q2, Q3, Q4) maîtrisent au moins 80 % des compétences et gardent ce niveau sur toute la période.

Les résultats de PISA 2012

Les résultats de PISA 2012 en compréhension de l'écrit indiquent un niveau moyen stable et au-dessus de la moyenne des pays de l'OCDE⁴.

Ce score est exactement celui obtenu en l'an 2000 mais se caractérise par une variabilité plus grande. En effet, la proportion d'élèves « très performants » (Correspondant aux niveaux 5 ou 6 sur l'échelle de compréhension de l'écrit développé par l'OCDE) est passée de 9 % à 13 %, et celle des élèves peu performants (en dessous du niveau 2, niveau seuil de compétence) est passée de 15 % à 19 %. En terme de score, les 10 % d'élèves les moins performants ont vu leurs résultats chuter de 23 points entre 2003 et 2012, alors que les 10 % d'élèves les plus performants ont progressé de 20 points ; l'écart entre ces deux groupes d'élèves s'est donc creusé de 43 points.

⁴ Voir OCDE (2013) : Note par pays. *In Programme international pour le suivi des acquis des élèves (PISA) résultat du PISA 2012*. Paris, OCDE. <http://www.oecd.org/pisa/keyfindings/PISA-2012-results-france.pdf>

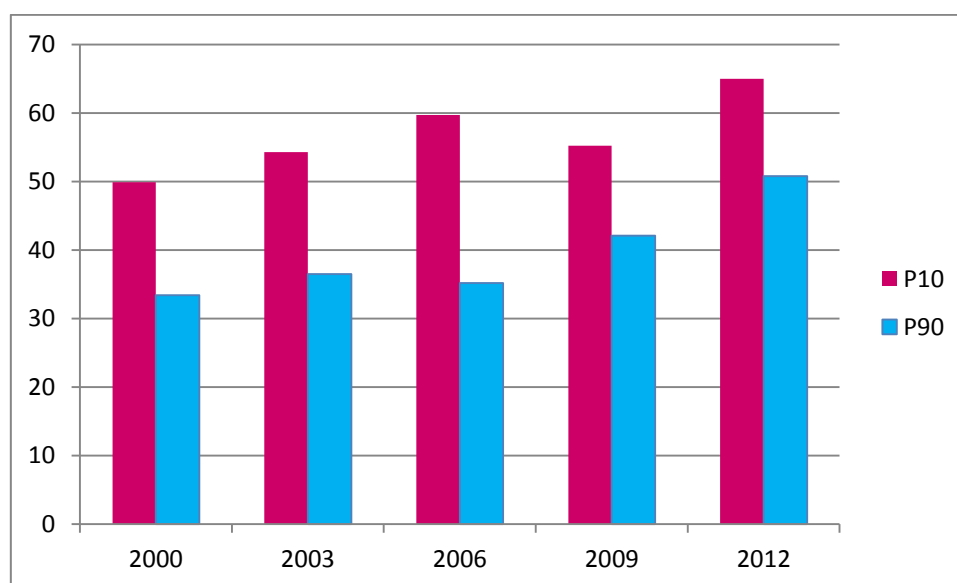
La France, avec une différence de score de 281 points – soit près de 3 écarts-type, présente l'écart de performance le plus important entre les 10 % de ses élèves les plus performants et les 10 % de ses élèves les moins performants, après Israël, où l'écart entre ces deux groupes d'élèves représente 295 points.

La compréhension de l'écrit des élèves de 15 ans scolarisés en France n'a donc pas vraiment évolué en moyenne en 10 ans, mais les écarts de performances se sont beaucoup accrus ; les bons élèves sont devenus meilleurs et les élèves en difficulté le sont encore davantage. Ces différences sont aussi très genrées. En effet, entre 2000 et 2012, la proportion d'élèves très performants a augmenté de 6 % chez les filles (contre seulement 2 % chez les garçons), alors que dans le même temps, la proportion d'élèves en difficulté a augmenté de 6 % chez les garçons (contre seulement 2 % chez les filles). Au final, en France, l'écart de performance en compréhension de l'écrit entre les sexes s'est creusé entre les cycles PISA 2000 et PISA 2012, passant de 29 à 44 points de différence en faveur des filles.

Enfin, l'écart de performance est également plus marqué par l'origine sociale que par le passé. La contribution de Monseur et Baye (2016, à paraître)⁵ montre que le poids de l'origine sociale sur les performances en lecture – mesuré par des régressions quantile des scores en lecture sur l'indice d'origine sociale – est plus marqué pour les élèves faibles et s'accroît sur la période.

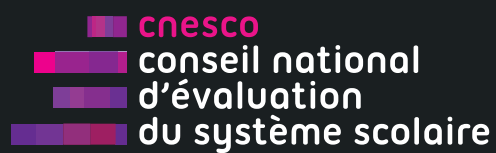
Ce résultat est probablement l'élément le plus saillant des enquêtes PISA, et concerne également les autres disciplines (mathématiques et culture scientifique).

Figure 1: Déterminisme social en France en lecture aux percentiles 10 et 90 conditionnellement à l'origine sociale, PISA 2000-2012, d'après Monseur et Baye (2016, à paraître)



Sources : enquêtes PISA, calculs Monseur et Baye

⁵ Monseur et Baye (2016, à paraître) : Quels apports des données PISA pour l'analyse des inégalités scolaires ? in Cnesco : *Les inégalités scolaires d'origine sociales et ethnoculturelle : une possible amplification ?* Paris, Cnesco.



Cnesco

Carré Suffren

31-35 rue de la Fédération

75 015 Paris

cnesco.communication@education.gouv.fr

École normale supérieure de Lyon
Institut français de l'Éducation

19 allée de Fontenay

69 007 Lyon

conf.consensus.ife@ens-lyon.fr