



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*

depp Direction de l'évaluation,
de la prospective
et de la performance

ÉTUDE DES PROGRESSIONS AU COLLÈGE À PARTIR DU PANEL 2011

Méthodologie psychométrique et modélisation économétrique

Marie-Camille Delarue, Laure Heidmann, Gaël Raffy

Série Etudes

Document de travail n° 2024-E25
Décembre 2024

ÉTUDE DES PROGRESSIONS AU COLLÈGE À PARTIR DU PANEL 2011

Méthodologie psychométrique et
modélisation économétrique



Direction de l'évaluation, de la prospective et de la performance

61-65, rue Dutot
75732 Paris Cedex 15

Directrice de la publication

Magda Tomasini

Auteurs

Marie-Camille Delarue, Laure Heidmann, Gaël Raffy

Série études : e-ISSN 2779-3532

SOMMAIRE



➤ Introduction.....	7
➤ Littérature	9
➤ I Données et estimation des compétences	10
A. Données.....	10
1. Évaluation de français et de mathématiques	10
2. Environnement familial	12
3. Repondération.....	15
B. Mesure des compétences des élèves	16
1. Méthodes d'estimation des scores des élèves.....	16
<i>Modèles de réponse à l'item.....</i>	<i>16</i>
<i>Procédures d'estimation</i>	<i>17</i>
<i>Calibration concurrente.....</i>	<i>18</i>
2. Analyses sur les items.....	19
<i>Fidélité du test</i>	<i>19</i>
<i>Validité des items.....</i>	<i>19</i>
<i>Regroupement d'items.....</i>	<i>20</i>
<i>Fonctionnement différentiel d'item.....</i>	<i>21</i>
3. Création d'une échelle commune	22
4. Ajustements apportés aux résultats.....	24
5. Tests de robustesse	24
➤ II Progression des élèves au collège.....	26
A. Progression moyenne et implications.....	26
1. Progression plus forte en mathématiques qu'en français.....	26
2. Comparaison des progressions avec les États-Unis.....	27
B. Différences d'apprentissage entre les élèves.....	28
➤ III Les facteurs de progression au collège.....	30
A. Résultats empiriques	31
1. Hétérogénéité des progressions selon le genre et le milieu scolaire au collège	33

2. Creusement des inégalités au collège	34
B. Répartition de la variance.....	35
C. Hypothèses du modèle	37
1. Linéarité	37
2. Homoscédasticité.....	37
3. Absence de multicolinéarité	38
4. Hypothèses propres au modèle à deux niveaux	39
D. Tests de robustesse	41
↘ Conclusion	43
Références des publications de la DEPP	45

↳ Résumé

Grâce aux données collectées auprès de 15 000 élèves suivis du CP à la troisième, cet article étudie l'évolution des apprentissages en français et en mathématiques au collège. Tout d'abord, nous utilisons des méthodes psychométriques pour estimer les compétences des élèves sur une échelle commune entre les différents niveaux scolaires, nous permettant ainsi de mesurer leur progression. Ensuite, nous étudions les facteurs influençant cette progression avec un modèle de croissance. Nous montrons que les élèves progressent plus rapidement en mathématiques qu'en français, avec une augmentation moyenne de score de 0,34 écart-type par an en mathématiques et de 0,16 en français. Au-delà de ces tendances moyennes, nous illustrons l'hétérogénéité des trajectoires. Nous observons que les filles connaissent de plus forts gains de compétence que les garçons au collège, aussi bien en français qu'en mathématiques. Nos résultats mettent également en évidence une aggravation des inégalités scolaires au collège, avec une progression plus marquée dans les milieux favorisés, et également dans les établissements privés.

STUDY OF MIDDLE SCHOOL PROGRESSIONS FROM THE 2011 PANEL

Psychometric methodology and econometric modeling

Abstract

Using data collected from 15,000 students surveyed from first grade to ninth, this article examines students' progress in middle school. First, we use psychometric methods to estimate students' skills on a common scale across grade levels, and thus measure their progression. We then study the factors involved in progression using a multilevel model. We show that students progress at a faster rate in mathematics than in French, with a mean score increase of 0.34 standard deviation per year in mathematics and 0.16 in French. Beyond these average trends, we illustrate the heterogeneity of trajectories. We observe that girls progress faster than boys in middle school, widening the gap in French while almost catching up with boys in mathematics. Lastly, our results highlight a worsening of educational inequalities at middle school, with a more marked progression in advantaged backgrounds, and also in private schools.

Traduction : auteur

Mots-clés : progression, collège, psychométrie, compétences, inégalités scolaires, panel

Keywords: Progression, middle school, psychometrics, skills, school inequalities, panel

➤ Introduction

Les résultats de PISA 2022 pour les élèves de 15 ans réitèrent une fois de plus le même constat : la France se positionne parmi les pays les plus inégalitaires selon l'environnement social des élèves en compréhension de l'écrit et en culture mathématiques (Bret *et al.*, 2023 ; Bernigole *et al.*, 2023). Pourquoi et comment en est-on arrivé à cette situation ? À la fin des années 1980, Duru-Bellat et Mingat estimaient que le collège produisait, dans ses deux premières années, plus d'inégalités sociales de résultats que toute la scolarité antérieure (Duru-Bellat & Mingat, 1993). L'hétérogénéité des parcours qui en résultent fait de ces années scolaires à la fois un défi et un terrain d'analyse. Que se passe-t-il tout au long du collège pour conduire à de telles disparités socio-économiques de résultats chez ces élèves de 15 ans ? Saisir l'évolution de leurs apprentissages afin de comprendre où se forment ces écarts représente un véritable enjeu éducatif. Et la mesure des progressions (plurielles) devient essentielle pour analyser les disparités dans l'acquisition des compétences au collège et identifier les principaux facteurs qui les influencent. Dès lors, une fois mises en lumière ces différences de parcours au collège, comment les expliquer ? Dans quelle mesure sont-elles imputables à l'origine sociale des élèves, aux dynamiques internes propres aux établissements scolaires, ou bien aux inégalités scolaires antérieures ? Le rôle du milieu social est souvent mis en cause, sous l'angle de la reproduction sociale, mais comment le collège endigue ou au contraire aggrave ce phénomène ? Toutefois, ce même collège accueille des élèves de niveau très inégal. Dans la mesure où les acquis scolaires sont cumulatifs, ces inégalités initiales (observées à l'école primaire) pourraient être la source de celles observées au collège.

Pour appréhender les mécanismes sous-jacents à ces disparités d'apprentissage, la mesure de la progression des élèves revêt de multiples intérêts. Tout d'abord, connaître le gain de compétence lié à une année scolaire permet de mieux comprendre les écarts entre niveaux : un bon élève de sixième est-il meilleur qu'un élève moyen de quatrième ? Cela permet également d'identifier les années où les élèves réalisent les plus grandes avancées dans leurs apprentissages. Aussi, mesurer l'hétérogénéité des compétences dans le temps offre une meilleure compréhension des différences de compétence entre les élèves. En effet, il sera alors possible d'interpréter les écarts de niveaux entre élèves en matière d'avance ou de retard en mois ou année d'apprentissage. De plus, ces mesures ouvrent une perspective de comparaison internationale, en mettant en regard le rythme de progression des collégiens en France avec celui de leurs homologues d'autres pays. Enfin, ces indicateurs de progression établis en conditions d'apprentissages normales pourront, le cas échéant, servir de référence pour interpréter les résultats d'interventions éducatives ou les conséquences d'une conjoncture particulière. Plus spécifiquement, ils pourront permettre de quantifier l'ampleur des effets d'une expérimentation en convertissant les gains du dispositif étudié en mois ou de semaines d'apprentissage.

Pour aborder ces différentes problématiques, ce document de travail et l'article associé (Delarue *et al.*, 2024) reposent sur le panel 2011, une enquête menée par la DEPP auprès de 15 000 élèves, suivis depuis leur entrée en CP en septembre 2011 jusqu'à la fin du collège. Lors de cette enquête, des données d'une grande richesse ont été collectées pour décrire les trajectoires de ces élèves : en plus du suivi des progrès avec chaque année des tests en français et en mathématiques, des questionnaires ont été adressés aux familles des élèves pour enrichir l'analyse en décrivant leur environnement quotidien.

La première contribution de notre étude est de mesurer la progression des élèves tout au long du collège. Pour y parvenir, un suivi longitudinal a été mis en place pour évaluer les élèves à différents moments de leur scolarité. L'un des défis majeurs est d'assurer la comparabilité des résultats au fil du temps. Grâce aux modèles de la théorie de réponse à l'item, nous estimons les compétences des élèves à différents niveaux scolaires sur une échelle commune. Cette approche nous permet de mesurer les progrès réalisés d'une année à l'autre et de comparer les compétences des élèves à

différents moments de leur scolarité. Au-delà de la progression moyenne, il existe une grande hétérogénéité dans les apprentissages des élèves. Notre deuxième contribution est d'intégrer des informations recueillies sur les élèves et leur environnement familial, afin de mesurer comment certains facteurs affectent les performances des élèves au cours du temps, en isolant chaque effet. Cette analyse repose sur des modèles de croissance, offrant une mesure précise des éléments qui façonnent les parcours scolaires.

Notre document de travail est structuré en plusieurs parties. Tout d'abord, une revue de la littérature situe le contexte de recherche dans lequel notre étude s'inscrit. Ensuite, la première partie présente les données et la méthode utilisées pour estimer les scores des élèves. La deuxième partie analyse les progrès des élèves et examine les types de trajectoires empruntées par les élèves tout au long du collège. Puis, la troisième partie identifie les influences de différents facteurs sur leurs apprentissages. Enfin, la dernière partie récapitule l'ensemble des résultats et conclut.

↳ Littérature

Notre étude se place dans la continuité de plusieurs travaux de recherche consacrés à l'évolution des apprentissages, afin d'actualiser les résultats et de les compléter grâce à l'opportunité offerte par la richesse des données collectées dans le cadre de ce suivi de panel.

De nombreuses études utilisent des tests standardisés pour mesurer le niveau des élèves, indépendamment des contextes spécifiques de chaque école et des pratiques de notation. Par exemple, dans une étude réalisée par Duru-Bellat et Mingat (1988), les progrès des élèves au collège ont été évalués en comparant leurs résultats aux évaluations de référence en CM2 avec des tests administrés en classe de cinquième, spécialement conçus pour les besoins de l'étude. Plus récemment, Cayouette-Remblière et Moulin (2019) se sont appuyés sur des évaluations diagnostiques en sixième et les notes obtenues aux épreuves du diplôme national du brevet (DNB) pour mesurer les progrès des élèves. Cependant, ces deux études se concentrent sur l'évolution relative du niveau des élèves, car il n'est pas possible de mesurer réellement les progressions en l'absence d'un référentiel commun entre les deux points de mesure. De plus, l'utilisation de deux modalités d'évaluation différentes peut introduire un biais lors de l'analyse des progressions relatives selon certains facteurs. Dans un article sur les parcours des filles et des garçons de l'école au supérieur, Chabanon et Jouveau (2022) montrent, en confrontant les écarts filles-garçons selon plusieurs modalités d'évaluation, que le diagnostic est sensible à ces modalités. Schématiquement, plus l'évaluation est de forme « scolaire » et interne à la classe, et donc s'éloigne d'une évaluation standardisée, plus les filles ont un avantage. Elles ont plutôt une meilleure réussite d'après les notes au contrôle continu du DNB que d'après les notes à l'examen terminal, et une meilleure réussite au DNB que lorsqu'on mobilise des évaluations standardisées. Par conséquent, une analyse des progressions des élèves à partir des évaluations de sixième et les notes au brevet risque de conclure à tort que les filles progressent plus que les garçons au collège, simplement parce que les modalités d'évaluation à la deuxième date (DNB) leur sont plus favorables qu'à la première.

Une mesure de la progression des élèves au collège a été réalisée par Grisay (1997) en français et en mathématiques. Pour obtenir des estimations des compétences comparables dans le temps, l'auteure se place dans le cadre des modèles de réponses à l'item à un paramètre (Rasch, 1980). Néanmoins, l'auteure souligne des problèmes qui ont limité la précision des résultats, notamment au moment de la conception des tests qui manquent d'items communs d'une année à l'autre pour assurer la comparabilité, ainsi que l'attrition élevée au cours de l'enquête (36 %) qui concerne particulièrement les élèves en difficulté. Une étude plus récente évalue la progression des élèves au collège, construite autour de multiples compétences comme, par exemple, les connaissances lexicales et la lecture silencieuse, à partir des données du panel de collégiens initié par la DEPP en

2007 (Ben Ali & Vourc'h, 2015). Les auteurs étudient le rôle de certains facteurs sociaux et familiaux sur la progression des élèves. Cependant, ils disposent de seulement deux points de mesure, un test en sixième et un autre en troisième.

Notre étude s'inscrit dans la lignée de ces travaux et bénéficie d'une collecte de données inédite. Tout d'abord, les élèves ont été évalués à cinq reprises, de la fin du CM2 jusqu'à la classe de troisième, en français et en mathématiques, ce qui permet d'obtenir une estimation précise de leur progression tout au long du collège. Des tests standardisés ont été spécialement conçus en respectant le cadre des méthodes psychométriques afin de garantir la robustesse et la comparabilité des résultats entre les différents niveaux scolaires. De plus, notre étude intègre les informations recueillies à partir de questionnaires adressés aux parents des élèves, ce qui enrichit considérablement le dispositif de recherche. Ces informations nous permettent de mieux comprendre les trajectoires individuelles des élèves et d'expliquer les facteurs qui peuvent influencer leurs progrès scolaires.

Enfin, notre article s'inscrit dans un contexte de recherche internationale qui vise à quantifier les gains d'apprentissage réalisés au cours d'une année scolaire. Plusieurs travaux de l'OCDE se sont intéressés à cette progression, en particulier chez les élèves de 15 ans dans le cadre du programme PISA. Leurs études les plus récentes, réalisées par Francesco Avvisati et Pauline Givord (Avvisati et Givord, 2021 ; Avvisati et Givord, 2023), mobilisent une méthodologie fondamentalement différente avec des données en coupe et non longitudinales : PISA ne suit pas les élèves à travers le temps. Dès lors, la progression est mesurée en comparant les scores à PISA d'élèves de niveaux scolaires consécutifs selon deux approches. Dans un cas, les auteurs utilisent les réglementations d'âge d'entrée à l'école qui font que selon son mois de naissance, un élève interrogé à PISA sera en général plus ou moins avancé dans son parcours scolaire ; dans un autre cas, les auteurs considèrent uniquement des pays qui ont changé leur mois d'interrogation dans PISA entre deux dates d'enquête. Ces deux sources exogènes entraînant un écart d'une année d'étude permettent l'estimation de l'effet d'une année scolaire par la méthode des doubles différences.

D'autres approches s'intéressant aux progressions des élèves ont été développées aux États-Unis mais cette fois-ci, avec l'objectif d'établir des repères empiriques pour mesurer la taille de l'effet d'interventions éducatives sur les résultats scolaires. Les travaux de Bloom, en particulier, quantifient dans cette optique les progrès réalisés par les élèves lors d'une année « normale » (Bloom *et al.*, 2008). Leur méthodologie s'appuie sur plusieurs tests nationaux standardisés avec des scores mis à l'échelle pouvant être comparés d'un niveau à l'autre, se rapprochant plus de nos travaux. Leurs résultats pourront ainsi servir de point de comparaison.

➤ I Données et estimation des compétences

A. Données

La scolarité de 15 188 élèves a été suivie pendant dix ans, depuis leur entrée en classe de CP en 2011, jusqu'au début du lycée, dans le cadre du panel 2011 (Caille, 2017). En plus d'évaluer les compétences chaque année, ce panel a pour ambition de croiser la progression des élèves avec des informations très riches sur leur environnement familial.

1. Évaluation de français et de mathématiques

Le premier objectif de l'étude est de mesurer le niveau des élèves chaque année afin d'évaluer leur progression. Pour cela, les élèves ont passé un test en français et un en mathématiques à chaque fin d'année scolaire à partir du CM2. Les mêmes évaluations ont été proposées avec un an de décalage aux élèves qui ont redoublé une fois (désignés par la suite comme redoublants). Au printemps 2020,

les évaluations de troisième, ainsi que celles de quatrième pour les élèves redoublants, ont été annulées en raison de la crise sanitaire. Pour pallier cet imprévu, le panel a été apparié avec les tests de positionnement de début de seconde, passés par l'ensemble des élèves entrant au lycée. Le tableau 1 présente le nombre d'élèves évalués à chaque niveau scolaire. Dans cet article, le champ d'étude sera restreint aux élèves qui ont passé au moins trois évaluations sur les cinq, c'est-à-dire à 12 904 élèves sur les 15 188 suivis en CP. Des pondérations permettent d'assurer la représentativité nationale de ces 12 904 élèves, avec le détail de la méthode dans la sous-partie A.3 ci-dessous. Le tableau 2 présente la répartition des élèves de ce champ selon différentes variables caractéristiques.

TABLEAU 1 • Nombre d'élèves évalués chaque année

Classe	Année scolaire (élèves « à l'heure »)*	Effectif	
		Total	Redoublant
CM2	2015-2016	13 957	988
6 ^e	2016-2017	12 449	882
5 ^e	2017-2018	11 203	967
4 ^e	2018-2019	11 078	annulé
3 ^e	2019-2020	annulé	annulé
2 ^{de}	2020-2021	11 220	782

Lecture : au total, 13 957 élèves ont passé l'évaluation du panel en CM2 dont 988 élèves redoublants.

*L'année scolaire des élèves redoublants est à lire avec un décalage d'une année. Ainsi, les redoublants de 2^{de} ont passé l'épreuve lors de l'année 2021-2022.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011.

Source : DEPP, panel 2011.

TABLEAU 2 • Répartition des élèves de notre champ

Variable	Modalités	Effectifs et proportions bruts		Proportions pondérées
		Effectifs	Proportions	
Sexe	Fille	6 474	50,2 %	49,3 %
	Garçon	6 430	49,8 %	50,7 %
Secteur de scolarisation en sixième ¹	Privé sous contrat	2 710	21,0 %	21,7 %
	Public hors EP	8 084	62,6 %	64,5 %
	REP	1 407	10,9 %	10,8 %
	REP+	703	5,4 %	3,0 %
Diplôme des parents	Sans diplôme ou brevet	1 313	10,2 %	9,7 %
	CAP ou BEP	2 208	17,1 %	17,2 %
	Baccalauréat	2 605	20,2 %	20,5 %
	Licence	2 716	21,0 %	21,4 %
	Master ou plus	3 142	24,3 %	24,3 %
	Non renseigné	920	7,1 %	7,0 %

Lecture : notre échantillon d'analyse comprend 6 474 filles, représentant 50,2 % de notre population totale et 49,3 % de notre population après pondération.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois tests de compétence et complété l'un des questionnaires familles.

Source : DEPP, panel 2011.

Les tests de français et de mathématiques évaluent plusieurs domaines de compétence. En français, des exercices de grammaire, d'orthographe, de compréhension écrite et de compréhension orale sont proposés. En mathématiques, les exercices portent notamment sur des compétences en nombres et calcul, en géométrie, en grandeurs et mesures et en gestion de données. Plusieurs exercices, appelés *items* par la suite, sont repris à l'identique d'une année sur l'autre. Ces items d'ancrage entre les tests permettent de comparer le niveau des élèves entre deux années, et sont essentiels pour mesurer leur progression.

¹ Le secteur du collège où sont scolarisés les élèves en sixième est utilisé dans l'ensemble de nos analyses. 4 % des élèves de notre échantillon ont changé de secteur entre la 6^e et la 4^e.

Deux expérimentations ont été mises en place à la rentrée 2022 pour créer un « pont » entre des évaluations qui n'avaient pas d'items communs. D'une part, un millier d'élèves de quatrième ont passé les tests de mathématiques de cinquième et de quatrième. D'autre part, un échantillon représentatif d'élèves de seconde a passé les tests de quatrième de français et de mathématiques, afin de pouvoir les comparer avec les tests de positionnement de seconde.

2. Environnement familial

Les familles des élèves du panel 2011 ont été interrogées deux fois, à quatre ans d'intervalle en 2012 (année du CP pour les non-redoublants) et 2016 (année du CM2). L'objectif est d'obtenir des informations sur l'environnement familial des élèves, à travers un grand nombre de questions portant notamment sur les ressources économiques des familles (les revenus et professions des parents, les conditions matérielles, etc.), ainsi que sur leurs habitudes culturelles (les activités extra-scolaires, le nombre de livres au domicile, etc.).

Afin de synthétiser la richesse des informations collectées dans ces questionnaires, des indices agrégés ont été construits selon certains axes d'intérêt. L'analyse des correspondances multiples (ACM) a d'abord permis d'identifier des grands axes du questionnaire, puis de construire des indices résumant les sous-ensembles de questions. Trois grandes dimensions se dégagent du questionnaire : la première est portée par l'environnement social, économique et culturel de la famille, la deuxième décrit les stratégies scolaires mises en œuvre par les parents, et la troisième résume l'implication personnelle des parents dans la scolarité de leurs enfants².

Nous introduisons dans la figure 1 la représentation graphique de l'ACM, où le premier axe correspond à notre indice caractérisant l'environnement socio-économique dans lequel baignent les élèves. Cette visualisation spatiale des axes de l'ACM projette les modalités des variables, permettant d'observer les corrélations entre les variables qui ont participé à la construction des axes et d'interpréter le sens de chaque dimension.

Les variables utilisées dans cette ACM sont :

- la profession et catégorie socioprofessionnelle (PCS) de la personne de référence du ménage, classée selon la nomenclature de l'Insee ;
- le plus haut diplôme de chacun des deux parents ;
- le nombre de livres chez soi (en tranches) ;
- le niveau de vie du ménage (calculé à partir du revenu par foyer et de la situation familiale) en quatre tranches délimitées par le seuil de pauvreté, le niveau de vie médian et un seuil de richesse (défini comme deux fois le niveau de vie médian) ;
- la pratique d'activités culturelles (bibliothèque - cinéma, cirque, théâtre - exposition) ;
- la fréquence de visionnage de la télévision ;
- l'origine étrangère ou non des parents (variable binaire prenant la valeur 1 si l'un des deux parents est d'origine étrangère).

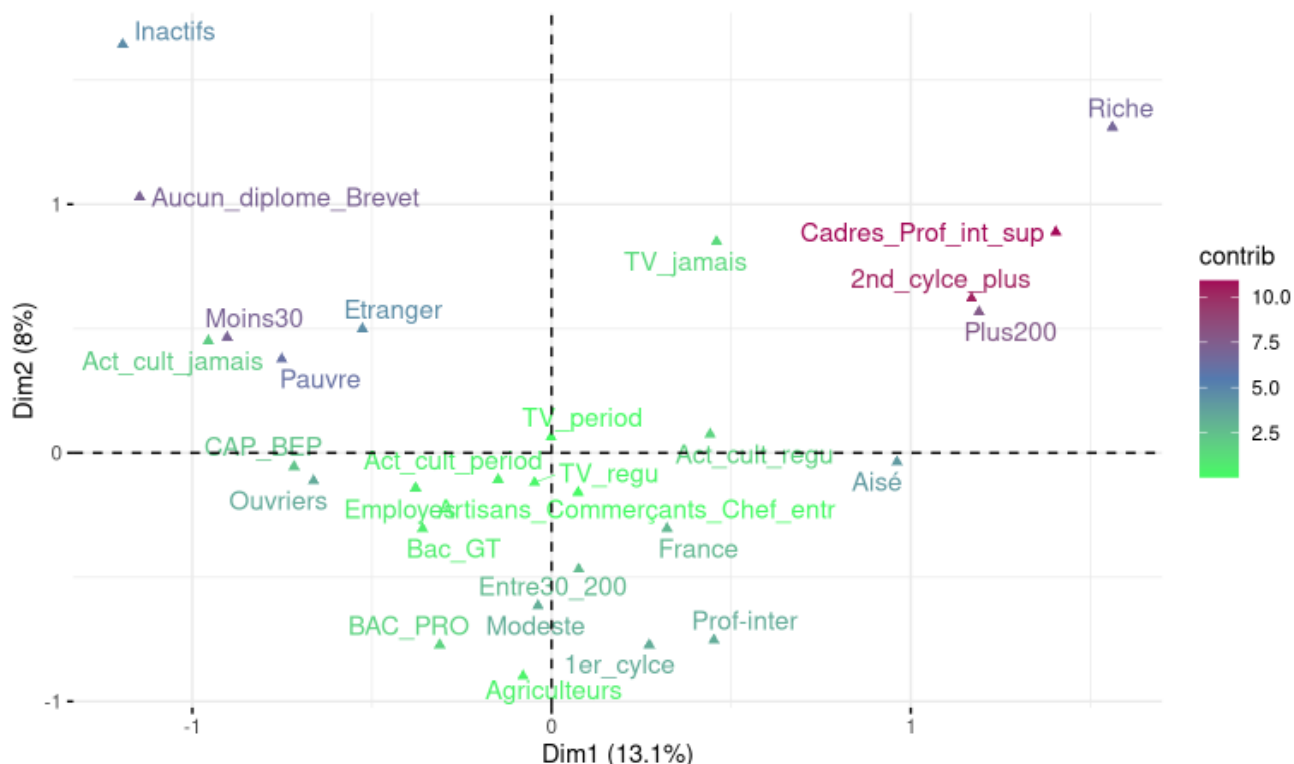
Les trois premières variables ci-dessus sont similaires à celles utilisées pour calculer l'indice ESCS dans PISA 2015. L'inclusion dans notre indice de variables supplémentaires vise à englober de nombreuses autres dimensions du capital social, économique et culturel des familles. Le choix de ces variables s'appuie sur des précédents établis dans d'autres études, telles que la construction de l'IPS (Rocher, 2016) utilisant le revenu, le visionnage de la télévision ou la pratique d'activités culturelles comme médiateurs des résultats scolaires. Aussi, la variable « nationalité étrangère » a déjà été mobilisée pour construire des indicateurs synthétiques - sous forme d'indices ou de typologies - pour caractériser le contexte socio-économique des territoires et établissements scolaires (Trancart, 2012). D'après les études qui ont considéré le lien entre l'origine étrangère d'un élève et ses performances scolaires,

² La construction de ces indices repose sur la méthodologie mise en place par l'OCDE dans le cadre de PISA 2015 pour calculer un indice du statut économique, social et culturel (ESCS), autour d'une analyse en composante principale (OECD, 2020).

cette dernière est un critère pertinent du contexte familial, car souvent associée à des performances scolaires plus faibles, y compris à autres variables d'environnement fixées (Ichou, 2013).

Cette analyse montre que des revenus élevés pour le foyer vont de pair avec des hauts diplômes des parents, un grand nombre de livres à la maison, des activités culturelles régulières et un faible visionnage de la télévision. Le premier axe oppose donc les familles aisées, dont les parents ont fait des études et exercent un métier valorisé socialement, aux familles issues de milieux défavorisés où les parents sont inactifs, sans diplôme, d'origine étrangère ou pratiquant peu d'activités culturelles.

FIGURE 1 • Projection sur les deux premiers axes de l'ACM des variables à trait socio-économique



Lecture : les modalités des variables avec un profil similaire sont regroupées. Celles corrélées négativement sont positionnées de part et d'autre de l'origine du graphique. L'échelle de couleur indique la contribution des modalités à la construction des axes.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois tests de compétence et complété l'un des questionnaires familles.

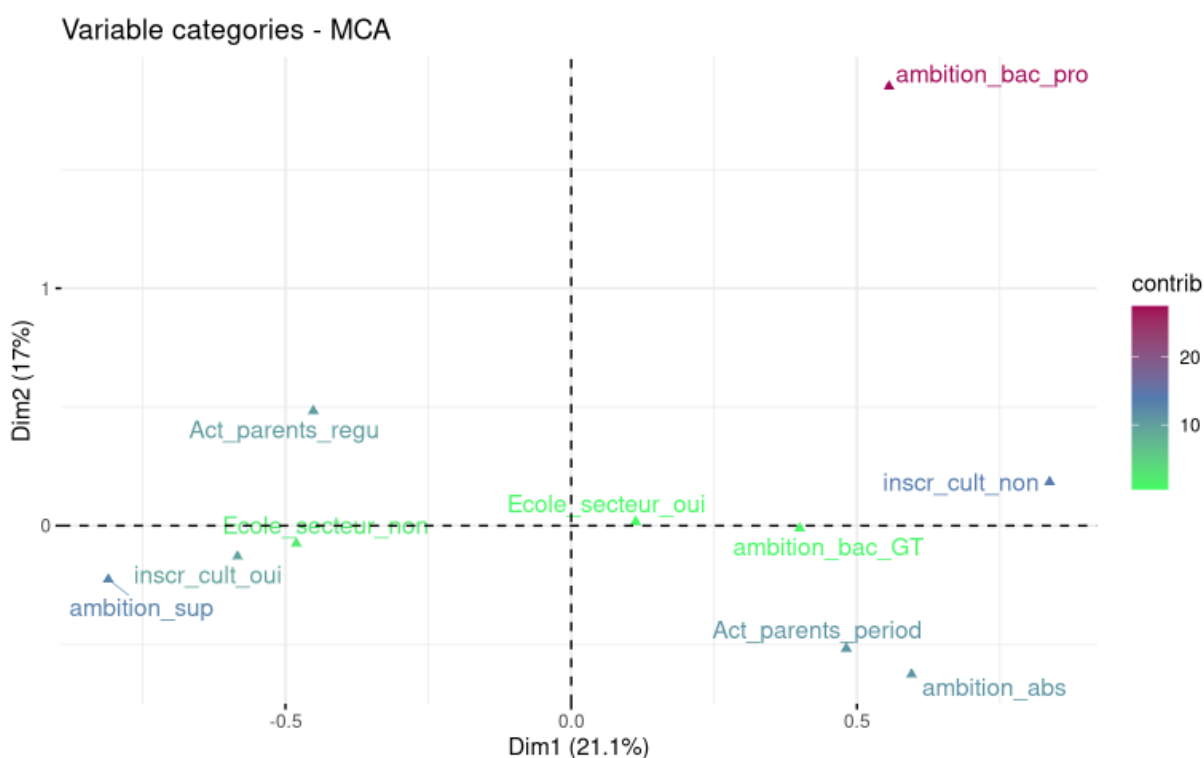
Source : DEPP, panel 2011.

Deux points sont à noter à la suite de cette analyse. Premièrement, la dimension sociale est fortement corrélée à la dimension économique, ce qui rend difficile leur séparation en deux indices distincts. D'autre part, la distribution des modalités présente une forme en « V » qui illustre l'effet Guttman (Cibois, 2014). Dans cette configuration, les variables extrêmes qui s'opposent fortement (à droite et à gauche) sont représentées vers le haut, tandis que les variables de position moyenne se situent au centre.

Enfin, il est intéressant de souligner la distinction entre notre indice et l'indice de position sociale (IPS), largement utilisé par les acteurs du système éducatif. L'IPS consiste en la projection des PCS sur le premier axe d'une ACM, elle-même élaborée à partir d'un ensemble de variables socio-économiques et culturelles liées à la réussite scolaire (Rocher, 2016). Ainsi, notre indice est construit de la même manière que ce score factoriel (sur lequel sont projetées les PCS). Cependant, la principale divergence, expliquant le choix de notre indice, réside dans l'utilisation de variables différentes lors de l'analyse factorielle. D'une part, nous avons souhaité exploiter la richesse des données du questionnaire famille, en utilisant les informations directement issues de notre panel d'élèves. Nous évitons ainsi un niveau d'abstraction supplémentaire issu du calage des PCS de nos données sur un indice construit à partir d'autres données. D'autre part, nous choisissons ici les variables employées, comme l'utilisation de la PCS dans la construction de l'indice.

Un deuxième indice définit les stratégies mises en œuvre par les parents en vue de créer un cadre favorable à la réussite scolaire de leur enfant. Il englobe des activités culturelles considérées comme légitimes, telles que l'inscription à la bibliothèque, les cours de langues, de musique ou de danse, en accord avec le concept développé par Bourdieu (Bourdieu et Passeron, 1964). Il inclut également la dérogation à la carte scolaire (entendue ici comme le fait de ne pas inscrire son enfant dans l'école la plus proche du domicile) et la pratique d'activités entre parents et enfants, telles que le sport, les travaux manuels ou les jeux de société. Enfin, une question sur l'ambition des parents pour la scolarité de leur enfant illustre leur volonté de mettre en place des pratiques propices à la réussite scolaire. La figure 2 représente l'ACM qui a servi à la construction de cet indice (premier axe).

FIGURE 2 • Projection sur les deux premiers axes de l'ACM des variables portant sur les stratégies scolaires



Lecture : l'indice de stratégie des parents correspond à l'inverse du premier axe. Ainsi, une valeur élevée de l'indice est associée à l'inscription à des activités culturelles, la pratique d'activités parent-enfant, le choix d'une école hors secteur et l'aspiration des parents à des études supérieures pour leur enfant.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois tests de compétence et complété l'un des questionnaires familles.

Source : DEPP, panel 2011.

Le dernier indice capture l'implication personnelle des parents dans la réussite scolaire de leur enfant et repose sur deux variables : l'aide aux devoirs apportée par différents membres de la famille et la rencontre des enseignants à l'initiative des parents. Cet indice correspond au premier axe de l'ACM présentée ci-dessous.

FIGURE 3 • Projection sur les deux premiers axes de l'ACM des variables définissant l'implication des parents



Lecture : l'indice d'implication des parents correspond à l'inverse du premier axe. Ainsi, une valeur élevée de l'indice est associée à une aide aux devoirs régulière et la rencontre des enseignants.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois tests de compétence et complété l'un des questionnaires familles.

Source : DEPP, panel 2011.

3. Repondération

Afin de garder un échantillon représentatif des élèves malgré l'attrition résultant de la non-réponse observée au cours des années, nous calculons de nouvelles pondérations. Nous explicitons ici la méthode utilisée. Tout d'abord, dans notre analyse, nous distinguons plusieurs champs d'étude (construits séparément pour chaque discipline) :

- Pour l'estimation des compétences des élèves, le champ d'étude correspond aux élèves présents aux tests du niveau scolaire étudié. En conséquence, nous calculons des pondérations transversales pour chaque année scolaire séparément.
- Pour l'analyse des progressions (statistiques descriptives et modèles de croissance), le champ d'étude correspond aux élèves qui ont participé à au moins trois évaluations entre le CM2 et la troisième, et dont les parents ont répondu à au moins un des questionnaires familles.

Pour chacun des champs d'étude, le même processus de repondération en deux étapes est employé. Dans un premier temps, nous modélisons la probabilité individuelle de réponse de l'ensemble des élèves initialement échantillonnés à l'aide d'une régression logistique. Les variables explicatives incluses dans le modèle regroupent toutes les caractéristiques disponibles sur les élèves. Dans un second temps, nous utilisons les probabilités estimées par notre modèle pour créer des groupes de réponses homogènes par la méthode des K-means, en appliquant la méthode Haziza-Beaumont (Haziza & Beaumont, 2007 ; Deroyon, 2017). Nous en récupérons la proportion de répondants dans chaque groupe, correspondant à nos nouvelles probabilités de réponse. Enfin, les poids finaux sont obtenus en multipliant les poids d'échantillonnage de chaque unité répondante par l'inverse de sa probabilité de réponse telle que nous venons de la calculer.

Nous avons également souhaité donner des statistiques concernant l'attrition résultant de la non-réponse, plus précisément concernant les élèves qui n'ont pas passé au moins trois évaluations sur

les cinq ou qui n'ont pas complété au moins un des deux questionnaires familles. Le tableau 3 présente la répartition de ces élèves hors de notre champ des analyses selon différentes variables caractéristiques. Ils sont plus représentés parmi les établissements prioritaires, notamment les REP+, et chez les redoublants que parmi les élèves de notre champ.

TABLEAU 3 • Répartition des élèves hors de notre champ

Variable	Modalités	Effectifs	Proportions
Sexe	Fille	1 026	43,5 %
	Garçon	1 333	56,5 %
Secteur de scolarisation en sixième	Privé sous contrat	347	19,8 %
	Public hors EP	952	54,3 %
	REP	261	14,9 %
	REP+	192	11,0 %
Année scolaire	À « l'heure »	1 670	70,9 %
	En retard	687	29,1 %

Lecture : parmi les élèves exclus de notre champ d'analyse, nous recensons 1 026 filles, ce qui représente 43,5 % des élèves hors champ. À noter que le nombre total d'élèves hors champ est de 2 359 élèves. Ce chiffre est légèrement supérieur à la différence entre l'effectif total d'origine (15 188 élèves) et les 12 904 élèves de notre champ car il intègre des élèves ayant réalisé moins de trois tests dans une discipline mais plus de trois tests dans l'autre. Aussi, la somme des effectifs est légèrement inférieure pour le secteur et l'année scolaire car des informations étaient manquantes pour certains élèves.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 n'ayant pas réalisé au moins trois tests de compétence ou n'ayant pas complété l'un des questionnaires familles.

Source : DEPP, panel 2011.

B. Mesure des compétences des élèves

Afin de pouvoir comparer les compétences des élèves entre eux, mais aussi les compétences des élèves entre les différents moments de la scolarité, il est essentiel de calculer des scores sur une échelle commune. Les méthodes d'analyses psychométriques s'appuient sur des fondements théoriques permettant de garantir cette comparabilité temporelle entre les tests (Rocher, 2013). L'étude se place donc dans le cadre des modèles de réponses à l'item : les compétences des élèves sont mises en regard avec les difficultés des items. Grâce aux items communs entre les tests de différents niveaux, il est possible de comparer les résultats des élèves aux différentes épreuves.

Les analyses psychométriques sont menées séparément sur chaque discipline, le français et les mathématiques.

1. Méthodes d'estimation des scores des élèves

Modèles de réponse à l'item

On considère n élèves ayant passé une évaluation composée de m items. On note Y_{ij} la réussite de l'élève i à l'item j :

$$Y_{ij} = \begin{cases} 1 & \text{si l'élève } i \text{ réussit l'item } j \\ 0 & \text{si l'élève } i \text{ échoue à l'item } j \end{cases}$$

Dans une approche classique, on peut calculer le score brut de l'élève au test défini comme le nombre d'items réussis par l'élève :

$$S_i = \sum_{j=1}^m Y_{ij}$$

Dans notre étude, nous utilisons les modèles de réponse à l'item (MRI) qui ont l'avantage, par rapport aux scores bruts, d'évaluer les compétences des élèves sur une échelle de niveau indépendante du test, et donc comparable entre des tests différents. Ces modèles positionnent les élèves sur une échelle de compétence en fonction de la difficulté des questions auxquelles ils ont répondu juste.

Dans le modèle de Birnbaum (1968), choisi pour cette étude, chaque item j est modélisé par deux paramètres :

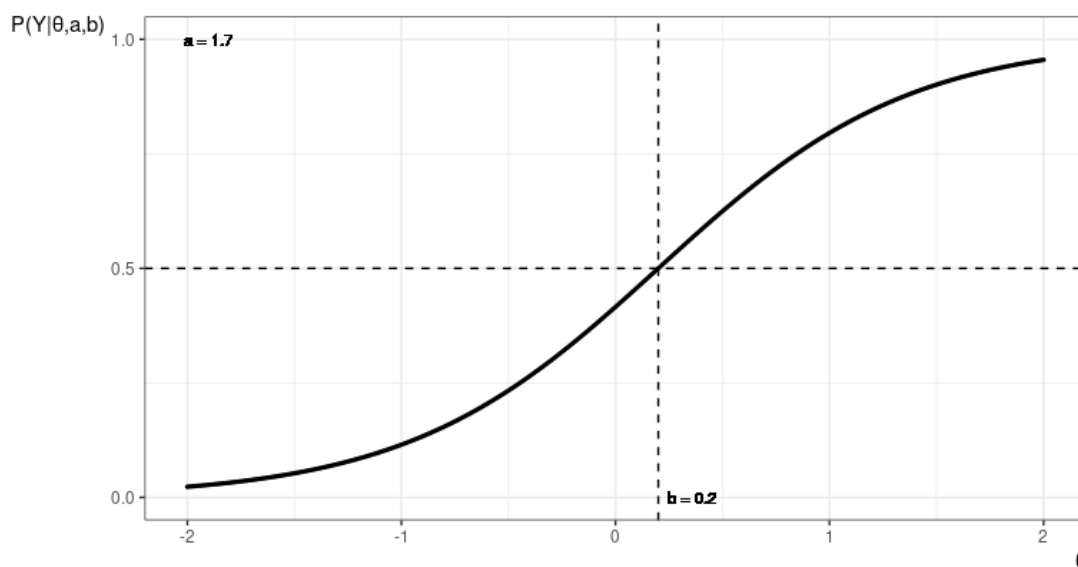
- la pente a_j qui indique sa capacité à discriminer les élèves selon leur niveau de compétence ;
- la difficulté b_j qui correspond au niveau de difficulté de la question.

Avec ces deux paramètres, il est possible de modéliser la probabilité qu'un élève i avec un niveau de compétence θ_i réponde juste à l'item j par la formule suivante :

$$P_{ij} = \mathbb{P}(Y_{ij} = 1 | \theta_i, a_j, b_j) = \frac{e^{a_j(\theta_i - b_j)}}{1 + e^{a_j(\theta_i - b_j)}}$$

L'objectif de l'analyse psychométrique est d'estimer, à partir des réponses des élèves, les paramètres de difficulté des items $(a_j, b_j)_{j \in [1, m]}$ ainsi que les compétences des élèves $(\theta_i)_{i \in [1, n]}$.

FIGURE 4 • Probabilité de réussir un item en fonction de la compétence de l'élève



Lecture : un élève avec un niveau de compétence de 0,2 à 50 % de chance de réussir cet item.

Procédures d'estimation

L'objectif est de trouver les paramètres et les compétences qui maximisent la vraisemblance d'observer les données observées. Cette vraisemblance correspond à la probabilité que les élèves, sachant leur niveau $(\theta_i)_{i \in [1, n]}$, répondent ce qu'ils ont répondu aux items de difficulté $(a_j, b_j)_{j \in [1, m]}$. Formellement, elle s'écrit :

$$\mathcal{L}(\theta, a, b) = \mathbb{P}((Y_1 = y_1) \cap \dots \cap (Y_n = y_n) | \theta, a, b)$$

Sous l'hypothèse que les élèves ne peuvent pas communiquer entre eux pendant les passations, les réponses au test sont indépendantes entre élèves. La vraisemblance peut alors s'écrire :

$$\mathcal{L}(\theta, a, b) = \prod_{i=1}^n \mathbb{P}(Y_i = y_i | \theta_i, a, b) = \prod_{i=1}^n \mathbb{P}((Y_{i1} = y_{i1}) \cap \dots \cap (Y_{im} = y_{im}) | \theta_i, a, b)$$

Sous l'hypothèse d'indépendance locale, qui stipule que la réussite à un item est indépendante à la réussite à un autre item, conditionnellement au niveau de l'élève, la vraisemblance se simplifie :

$$\mathcal{L}(\theta, a, b) = \prod_{i=1}^n \prod_{j=1}^m \mathbb{P}(Y_{ij} = y_{ij} | \theta_i, a_j, b_j) = \prod_{i=1}^n \prod_{j=1}^m \mathbb{P}(Y_{ij} = 1 | \theta_i, a_j, b_j)^{y_{ij}} \left(1 - \mathbb{P}(Y_{ij} = 1 | \theta_i, a_j, b_j)\right)^{1-y_{ij}}$$

En utilisant les modèles de réponses à l'item, la vraisemblance s'écrit :

$$\mathcal{L}(\theta, a, b) = \prod_{i=1}^n \prod_{j=1}^m \left(\frac{e^{a_j(\theta_i - b_j)}}{1 + e^{a_j(\theta_i - b_j)}} \right)^{y_{ij}} \left(1 - \frac{e^{a_j(\theta_i - b_j)}}{1 + e^{a_j(\theta_i - b_j)}} \right)^{1-y_{ij}}$$

Nous cherchons les paramètres et les compétences qui maximisent cette vraisemblance. Compte tenu de la complexité du problème, l'estimation est conduite en deux temps : l'estimation des paramètres des items, puis celle des compétences en considérant les paramètres des items comme fixes.

Pour estimer les paramètres d'item sans connaître les compétences des élèves, nous nous intéressons à la vraisemblance marginale : il n'y a plus besoin de connaître la compétence de chaque élève, mais seulement la distribution f des compétences des élèves dans la population. La vraisemblance est marginale dans le sens où les paramètres concernant les individus n'apparaissent plus dans la formule de vraisemblance :

$$\mathcal{L}(a, b) = \prod_{i=1}^n \prod_{j=1}^m \int_{-\infty}^{+\infty} \mathbb{P}(Y_{ij} = y_{ij} | \theta, a_j, b_j) f(\theta) d\theta$$

Classiquement, la distribution des scores des élèves est supposée suivre une loi normale centrée et réduite.

La résolution de ce problème de maximisation est effectuée à l'aide de l'algorithme EM (Expectation-Maximization), qui approxime les intégrales en utilisant des points de quadrature (Bock et Aitkin, 1981). L'algorithme se déroule en plusieurs étapes de manière itérative. Les valeurs des paramètres sont tout d'abord initialisées. Puis l'espérance conditionnelle de la vraisemblance des données est calculée avec les valeurs des paramètres estimées à l'étape précédente. Ensuite, cette espérance conditionnelle est maximisée pour trouver de nouvelles valeurs de paramètres. L'algorithme se poursuit ainsi de suite jusqu'à ce qu'un critère de convergence des paramètres soit atteint. Pour plus de détails, il est possible de se référer à Rocher (2013).

Une fois les paramètres des items estimés, ils sont considérés comme fixes et il est possible d'estimer les scores des élèves. En pratique, nous cherchons les $(\theta_i)_{i \in [1, n]}$ qui maximisent la vraisemblance $\mathcal{L}(\theta, a, b)$ avec $(a, b)_{j \in [1, m]}$ connus. Cependant, comme le nombre de paramètres à estimer, les $(\theta_i)_{i \in [1, n]}$, augmente avec le nombre d'observations, les propriétés classiques de l'estimateur du maximum de vraisemblance ne sont pas vérifiées. Il existe alors un biais noté $\mathcal{B}(\hat{\theta})$. Pour obtenir un estimateur non biaisé, Warm (1989) a proposé de maximiser une vraisemblance pondérée $w(\theta)\mathcal{L}(\theta, a, b)$, en choisissant $w(\theta)$ de manière à ce que l'annulation de la dérivée du logarithme de la vraisemblance pondérée revienne à résoudre l'équation suivante :

$$\frac{\partial \ln \mathcal{L}}{\partial \theta} + \mathcal{B}(\hat{\theta}) = 0$$

Grâce à ces deux procédures successives, la maximisation de la vraisemblance marginale puis la maximisation de la vraisemblance pondérée, nous pouvons estimer les scores des élèves en fonction de leurs réponses au test.

Calibration concurrente

Les modèles psychométriques offrent un avantage majeur : la capacité d'évaluer les compétences des élèves sur une échelle commune, même s'ils ont passé des tests différents. Pour cela, il est essentiel d'avoir des items en commun entre les tests, garantissant ainsi la comparabilité des niveaux de difficulté des questions.

Plusieurs méthodes existent pour établir cette comparabilité. Dans notre étude, nous avons opté pour la calibration concurrente. Cette méthode permet de rassembler plusieurs tests et d'estimer leurs paramètres d'items en une seule fois (Lee & Lee, 2018). Les données provenant de différents tests sont combinées pour former un grand jeu de données. Ensuite, l'estimation des paramètres est réalisée de manière à définir l'un des groupes (dans notre cas, le niveau scolaire le plus bas) comme groupe de référence, ayant une moyenne des scores fixée à 0 et un écart-type fixé à 1. Les moyennes

et écarts-types des scores des autres groupes peuvent différer de cette échelle de référence et sont estimés par rapport à elle. La calibration concurrente produit une seule estimation des paramètres pour les items communs. Grâce à cette calibration, toutes les estimations des paramètres des items des différents tests sont placées sur la même échelle de référence, sans nécessiter d'autres transformations. Une fois que les paramètres d'items ont été calculés sur cette échelle commune, il est possible d'évaluer les compétences des élèves en fonction de leurs réponses à chaque test, ces estimations étant automatiquement alignées sur l'échelle de référence.

2. Analyses sur les items

Avant de procéder aux estimations, nous effectuons quelques étapes de vérification des hypothèses du modèle afin de s'assurer de la validité des tests et de la possibilité de créer une échelle de compétence unique entre tous les niveaux.

Deux points sont à noter concernant l'élaboration des tests. D'une part, l'ensemble des items ont été répartis aléatoirement entre les élèves et entre les différentes épreuves afin d'éviter certains biais potentiels comme l'effet de fatigue. D'autre part, pour les items communs entre les tests, nous faisons l'hypothèse de l'absence de mémoire des élèves d'une année à l'autre.

Fidélité du test

Nous commençons par calculer la fidélité de chaque test du CM2 à la seconde, français et mathématiques séparément. La fidélité d'un test indique la précision avec laquelle le test évalue réellement la compétence qu'il prétend mesurer. Nous nous appuyons sur l'omega de McDonald (Béland et al., 2017), mesurant la cohérence interne des résultats obtenus. Nous considérons que la fidélité du test est satisfaisante lorsque la valeur du coefficient omega est supérieure à 0,80, ce qui est vérifié pour l'ensemble des tests comme on peut le voir dans le tableau 4. À noter qu'en seconde les tests sont adaptatifs (i.e. les élèves sont orientés vers des questions plus ou moins difficiles selon leur niveau), les coefficients omega sont alors calculés seulement sur les items qui sont passés par tous les élèves, ce qui peut expliquer leurs valeurs sensiblement inférieures à celles des autres tests.

TABLEAU 4 • Fidélité des tests

Classe	Français	Mathématiques
CM2	0,98	0,92
6 ^e	0,94	0,95
5 ^e	0,91	0,93
5 ^e – 4 ^e	-	0,92
4 ^e	0,90	0,87
4 ^e – 2 ^{de}	0,94	0,92
2 ^{de}	0,83	0,83

Lecture : l'omega de McDonald du test de mathématiques de cinquième vaut 0,93.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011.

Source : DEPP, panel 2011.

Validité des items

Ensuite, nous vérifions la validité des items, l'objectif étant d'écarter les questions qui mesurent mal la compétence visée. Le r biserial point (DeMars, 2018), qu'on nomme plus simplement r-bis, permet d'évaluer à quel point un item mesure la même compétence que les autres questions (Rocher, 2013). Plus précisément, c'est le coefficient de corrélation entre le vecteur $Y_j = (Y_{ij})_{i \in [1, n]}$ des variables indicatrices de réussite des élèves à l'item j et $S = (S_i)_{i \in [1, n]}$ le vecteur des scores bruts des élèves sur l'ensemble du test :

$$r_j^{bis} = corr(Y_j, S) = corr\left(Y_j, \frac{1}{m} \sum_{k=1}^m Y_k\right)$$

En pratique, nous utilisons le r-bis corrigé, qui exclut l'item en question du score global :

$$r_j^{bis\ corrigé} = corr(Y_j, S_{-j}) = corr\left(Y_j, \frac{1}{m-1} \sum_{k \neq j} Y_k\right)$$

Par la suite, les items dont la valeur du r-bis est inférieure à 0,20 sont considérés comme mesurant mal la compétence visée et sont alors supprimés des tests. Les élèves ne sont donc pas évalués sur ces questions. Les tableaux 5 et 6 indiquent le nombre d'items qui ont été supprimés en raison d'un r-bis trop faible. Dans ces tableaux, nous présentons aussi les analyses pour les expérimentations qui ont été rajoutées en 2022 entre la cinquième et la quatrième (seulement en mathématiques) et entre la quatrième et la seconde. Les deux dernières colonnes sont expliquées dans la section suivante.

TABLEAU 5 • Nombre d'items en français

Classe	Items	Items supprimés	Items d'ancrage (après suppression)	Items en FDI
CM2	199	3	-	-
6 ^e	88	4	28	13
5 ^e	67	8	28	0
4 ^e	55	2	10	0
4 ^e – 2 ^{de}	124	8	49	5
2 ^{de}	121	14	30	2

Lecture : le test de français de sixième comporte 88 items, dont 4 ont été retirés des analyses à cause d'une valeur de r-bis trop faible. Parmi les 84 items restants, 28 sont des items communs avec le test de CM2 pour constituer l'ancrage entre les deux tests. 13 items parmi ces 28 sont en fonctionnement différentiel et donc sont retirés de l'ancrage, mais pas des analyses.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011.

Source : DEPP, panel 2011.

TABLEAU 6 • Nombre d'items en mathématiques

Classe	Items	Items supprimés	Items d'ancrage (après suppression)	Items en FDI
CM2	49	1	-	-
6 ^e	95	6	32	13
5 ^e	46	0	26	1
5 ^e – 4 ^e	53	3	26	1
4 ^e	26	1	12	5
4 ^e – 2 ^{de}	90	10	15	4
2 ^{de}	115	21	58	11

Lecture : le test de mathématiques de quatrième comporte 26 items, dont 1 a été retiré des analyses à cause d'une valeur de r-bis trop faible. Parmi les 25 items restants, 12 sont des items communs avec le test de l'expérimentation pour faire le pont entre cinquième et quatrième pour constituer l'ancrage. 5 items parmi ces 12 sont en fonctionnement différentiel et donc sont retirés de l'ancrage, mais pas des analyses.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011.

Source : DEPP, panel 2011.

Regroupement d'items

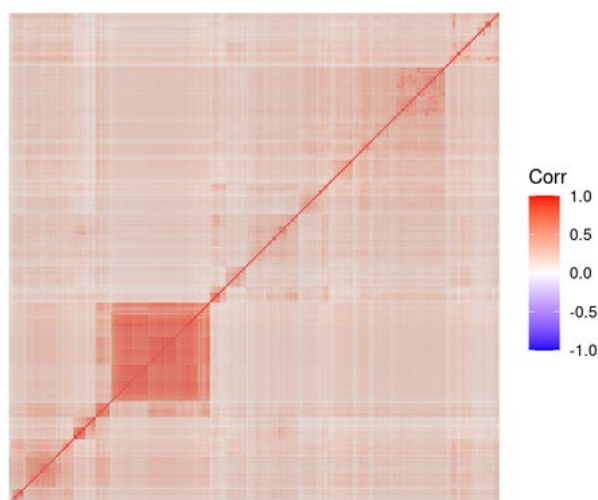
Les groupes d'items qui sont trop corrélés entre eux peuvent entraîner des problèmes dans l'estimation des compétences des élèves, en conduisant à une inflation artificielle dans la variance des scores. Pour éviter ce phénomène, nous regroupons les items qui sont trop corrélés entre eux en un item unique.

Pour chaque niveau scolaire, sur l'ensemble des évaluations de français et de mathématiques, la matrice des corrélations entre les réussites aux items est calculée. Les corrélations supérieures à 0,7 sont examinées en détail. Dans l'évaluation de français de CM2, nous repérons un groupe de 48 items très corrélés entre eux, présentés dans la figure 5. Nous les regroupons alors dans un item unique

dont le score final est déterminé selon le nombre de réponses justes : le score de 1 est attribué aux élèves qui ont réussi au moins 41 sous-questions parmi les 48.

Pour les autres évaluations, des petits groupes d'items fortement corrélés ont été repérés. Ce sont en général trois ou quatre items situés sur la même page, ou portant exactement sur la même sous-compétence. Par exemple, en français, deux items interrogent la conjugaison d'un verbe du premier groupe au présent, et la réussite à l'un est donc fortement corrélée à la réussite à l'autre. Des regroupements de ces items avaient été effectués en première analyse. Cependant, dans l'objectif de simplicité, et comme les scores n'étaient pas significativement affectés par le regroupement d'items, il a été finalement choisi de ne pas regrouper les items corrélés entre eux, hormis le groupe d'items de CM2 présenté plus haut.

FIGURE 5 • Matrice des corrélations des réussites aux items des tests de CM2



Lecture : chaque cellule de coordonnée (i, j) représente le coefficient de corrélation entre la réussite à l'item i (en ligne) et la réussite à l'item j (en colonne). Le grand carré rouge correspond à un groupe de 48 items de français très fortement corrélés entre eux.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011.

Source : DEPP, panel 2011.

Fonctionnement différentiel d'item

Afin de pouvoir mesurer la progression du niveau des élèves au cours de leur scolarité, des items, appelés items d'ancrage, sont proposés à l'identique d'une année sur l'autre. L'évolution de la réussite des élèves à ces questions permet de mesurer leur progression.

Toutefois, il est possible que la réussite à un item d'ancrage varie sans que cela traduise une baisse ou une hausse du niveau des élèves. Par exemple, si un item porte sur un sujet qui est médiatisé une année donnée, sa difficulté peut potentiellement diminuer sans correspondre à une hausse du niveau des élèves. On ne peut pas considérer que la difficulté de cet item est constante au cours du temps, puisque les élèves ont été davantage familiarisés avec l'item à un moment donné. Par conséquent, cet item ne peut plus servir de référence pour étudier l'évolution du niveau des élèves. On dit que l'item présente un fonctionnement différentiel.

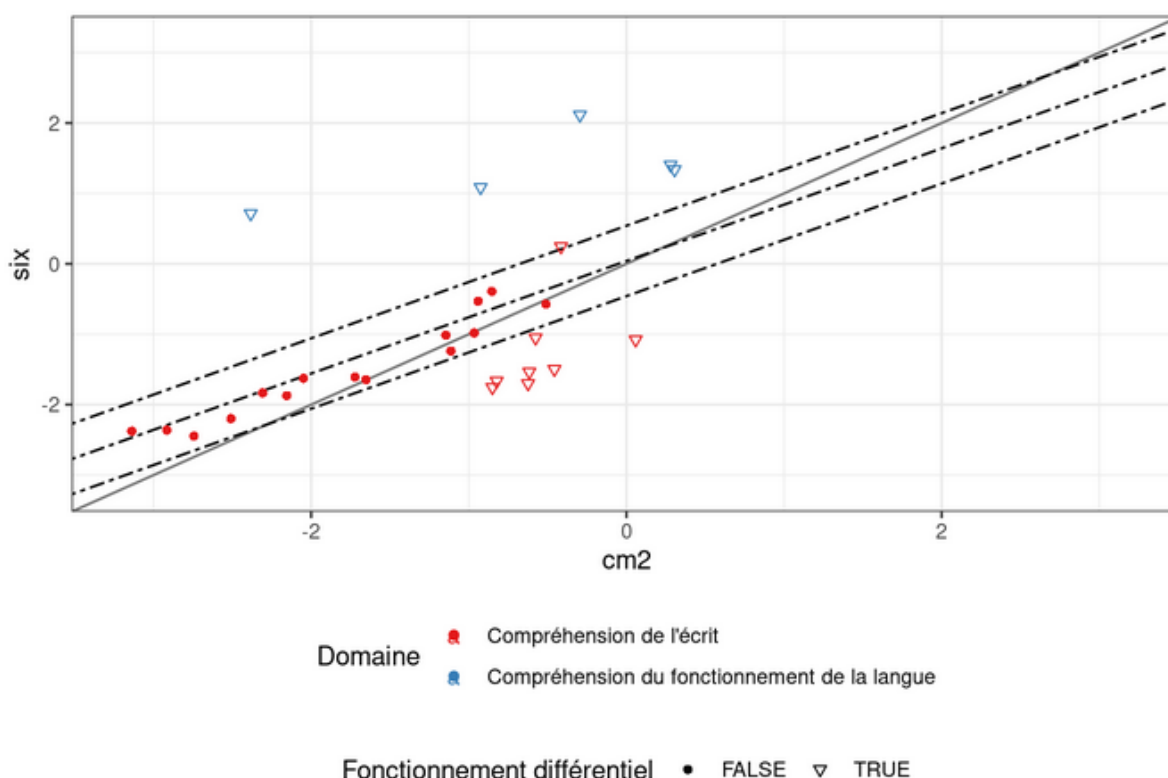
Dans notre étude, un item est considéré en fonctionnement différentiel lorsque l'évolution de ses paramètres entre deux années est trop différente de la tendance des évolutions des autres items. Pour repérer ces items, nous calculons leurs paramètres pour chaque évaluation à chaque niveau séparément. Puis, nous comparons les paramètres obtenus pour les items d'ancrage entre les différents niveaux scolaires. Il est alors possible de tracer la tendance de l'évolution des paramètres d'item entre chaque niveau. Les items qui s'éloignent trop de cette tendance sont en fonctionnement différentiel. Nous choisissons deux critères pour détecter un fonctionnement différentiel d'item (FDI) :

1. la difficulté de l’item mis sur l’échelle de l’année précédente s’écarte de sa valeur de l’année précédente de plus de 0,5 ;
2. ou la pente de l’item mis sur l’échelle de l’année précédente s’écarte de sa valeur de l’année précédente de plus de 1.

À la suite des analyses, beaucoup d’items communs entre le CM2 et la sixième présentent un fonctionnement différentiel, probablement dû à la différence de modalité de passation des épreuves, entre le papier pour le CM2 et l’ordinateur pour la sixième. En particulier, les réponses à certains items de français ne doivent pas simplement être cochées comme la plupart des items, mais doivent être écrites à la main en CM2 et tapées sur le clavier en sixième, ce qui correspond aux items en bleu dans la figure 6. La difficulté supplémentaire induite par le fait de taper à l’ordinateur est certainement à l’origine de la hausse de difficultés observées pour ces items en sixième. Ces items sont donc considérés comme différents entre le CM2 et la sixième puisqu’ils ne mesurent pas exactement la même compétence.

Les tableaux 5 et 6 présentent le nombre d’items d’ancrage entre deux niveaux scolaires, après avoir supprimé les items avec un r-bis trop faible. La dernière colonne indique le nombre de fonctionnements différentiels d’item. Ces items ne sont pas supprimés des données mais leurs noms sont modifiés pour les considérer comme des items différents entre les évaluations.

FIGURE 6 • Évolution des difficultés des items de français entre le CM2 et la sixième



Lecture : les items de compréhension de fonctionnement de la langue, en bleu dans la figure, ont une difficulté supérieure en sixième (valeurs en ordonnées) qu'en CM2 (en abscisse).

Champ : panel d'élèves de France hors DROM entrés en CP en 2011.

Source : DEPP, panel 2011.

3. Création d'une échelle commune

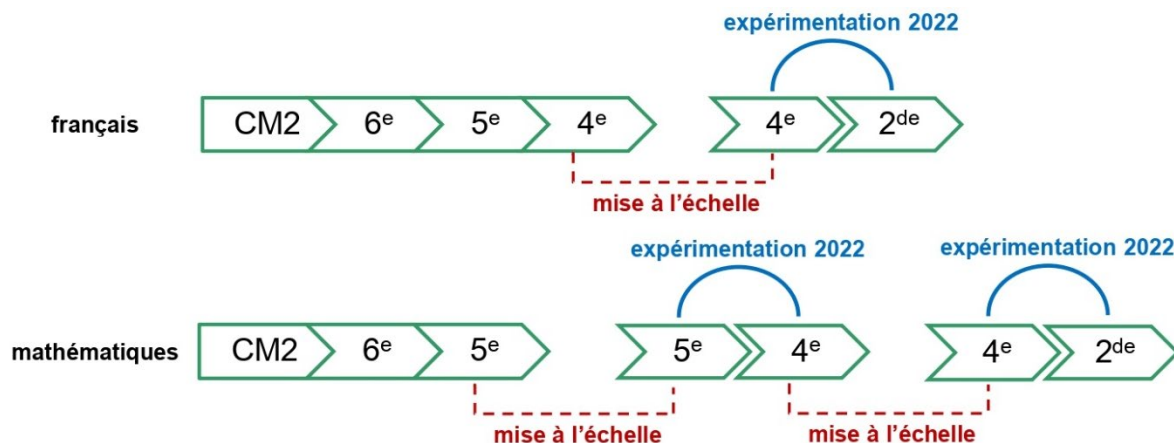
Les scores de français sont estimés en calibration concurrente en deux temps :

1. du CM2 à la quatrième ;
2. de la quatrième à la seconde en passant par l'expérimentation 4^e – 2^{de} de 2022.

Les scores de mathématiques sont estimés en calibration concurrente en trois temps :

1. du CM2 à la cinquième ;
2. de la cinquième à la quatrième en passant par l'expérimentation 5^e – 4^e de 2022 ;
3. de la quatrième à la seconde en passant par l'expérimentation 4^e – 2^{de} de 2022.

FIGURE 7 • Création des échelles de compétences en français et en mathématiques



Lecture : les compétences des élèves en français sont estimées en calibration concurrente du CM2 à la quatrième. Puis, une seconde calibration concurrente est effectuée entre la quatrième et la seconde, en passant par l'expérimentation de 2022. Cette seconde estimation est ensuite mise sur l'échelle de la première.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011.

Source : DEPP, panel 2011.

Comme schématisé dans la figure 7, nous disposons de deux échelles de scores en français et trois en mathématiques (une pour chaque calibration concurrente), où les scores aux tests du niveau scolaire de référence (CM2 pour le premier groupe, cinquième pour le deuxième et quatrième pour le dernier) sont distribués selon la loi normale centrée et réduite. Afin de créer une échelle commune par discipline, nous effectuons des transformations linéaires pour mettre ces échelles bout à bout. Ces transformations linéaires n'ont pas d'effet sur la validité des résultats car le problème de maximisation de vraisemblance n'est pas identifiable, *i.e.* il existe une infinité de solutions équivalentes à une transformation linéaire près (Rocher, 2013).

Par exemple, pour le français, nous effectuons une transformation linéaire du type $\theta \rightarrow A\theta + B$ pour ramener les scores de seconde sur l'échelle des scores de CM2. Pour déterminer les coefficients A et B , il faut comparer les scores de quatrième calculés dans la seconde estimation, avec les scores de quatrième obtenus sur l'échelle des scores de CM2. On note (μ^4, σ^4) la moyenne et l'écart-type de la distribution des scores de quatrième qui sont calculés dans la seconde calibration concurrente, et $(\mu_{ref}^4, \sigma_{ref}^4)$ sont la moyenne et l'écart-type des mêmes scores mais qui ont été obtenus lors de la première calibration concurrente. Pour déplacer les scores de quatrième sur l'échelle de référence, nous effectuons la transformation suivante :

$$\theta \rightarrow \sigma_{ref}^4 \times \frac{\theta - \mu^4}{\sigma^4} + \mu_{ref}^4$$

Les coefficients de conversion pour recoller l'échelle des scores de quatrième sur celle des scores de CM2 s'écrivent alors :

$$\begin{cases} A = \frac{\sigma_{ref}^4}{\sigma^4} \\ B = \mu_{ref}^4 - A\mu^4 \end{cases}$$

Une fois ces deux coefficients obtenus, les scores des élèves en seconde peuvent être rendus comparables avec ceux des niveaux précédents, en appliquant la transformation linéaire suivante : $\theta \rightarrow A\theta + B$.

4. Ajustements apportés aux résultats

Plusieurs évolutions du contenu de l'évaluation et du protocole de collecte entre les différents niveaux du panel ont eu un impact sur les résultats obtenus.

Tout d'abord, il est important de souligner que, contrairement aux autres évaluations, les tests de seconde ont été passés en début d'année scolaire. Dans cette étude, nous considérons qu'ils mesurent le niveau des élèves en fin de troisième, ce qui implique que les élèves gardent les mêmes compétences entre la fin d'année et le début d'année scolaire suivante. Or, certaines études suggèrent que les compétences scolaires des élèves diminuent pendant les vacances. Et cette baisse est plus prononcée chez les élèves de catégories sociales défavorisées, entraînant alors une accentuation des écarts de performance (Andrieux *et al.*, 2002). Le niveau des élèves à la fin du collège est alors potentiellement sous-estimé, en particulier pour les élèves défavorisés, à cause des deux mois de vacances d'été.

Les modalités de passations ont changé à l'entrée au collège : en CM2, les élèves répondent aux questions à l'écrit sur un cahier, alors qu'au collège ils passent les tests sur ordinateur, ce qui ajoute une difficulté supplémentaire puisque les élèves sont moins habitués à travailler sur support numérique. Une étude menée à la DEPP sur une évaluation Cedre (Cycle des évaluations disciplinaires réalisées sur échantillon) compare les résultats à un test de mathématiques passé alternativement sur cahier et sur ordinateur par des élèves de troisième, et conclut à un « effet papier » de 0,2 écart-type de score en plus pour les tests passés sur cahier (Dos Santos *et al.*, 2023). De plus, les auteurs constatent que cet effet est globalement homogène pour l'ensemble des élèves, même s'ils notent que les performances des filles sont très légèrement moins diminuées sur ordinateur. Cet effet se retrouve également en français. Pirls, une étude internationale qui évalue tous les cinq ans le niveau de compréhension de l'écrit des élèves de CM1, a réalisé en 2021 une étude dite de « bridge », visant à faire passer certains items aux deux formats, numérique ou papier (Yin, L. *et al.*, 2023). Si les auteurs n'ont pas quantifié d'effet en matière d'écart-type de score (soulignant leur but de relier les données numériques à l'échelle de Pirls et non pas de considérer des indicateurs de différence de mode), cette étude a pu donner des tendances significatives. Les pourcentages moyens de réponses correctes sont supérieurs de 6 % en moyenne sur papier. Par conséquent, nous retranchons un effet papier de 0,2 écart-type aux scores de CM2, aussi bien en français qu'en mathématiques, pour rendre comparables les scores malgré le changement de modalité de passation. Le choix de retrancher le même écart-type de 0,2 en français permet d'obtenir une progression entre le CM2 et la 6^e cohérente avec celles entre les autres niveaux scolaires. À noter que cette correction intervient après l'estimation par calibration concurrente et n'affecte donc pas le calcul des progressions entre les autres niveaux.

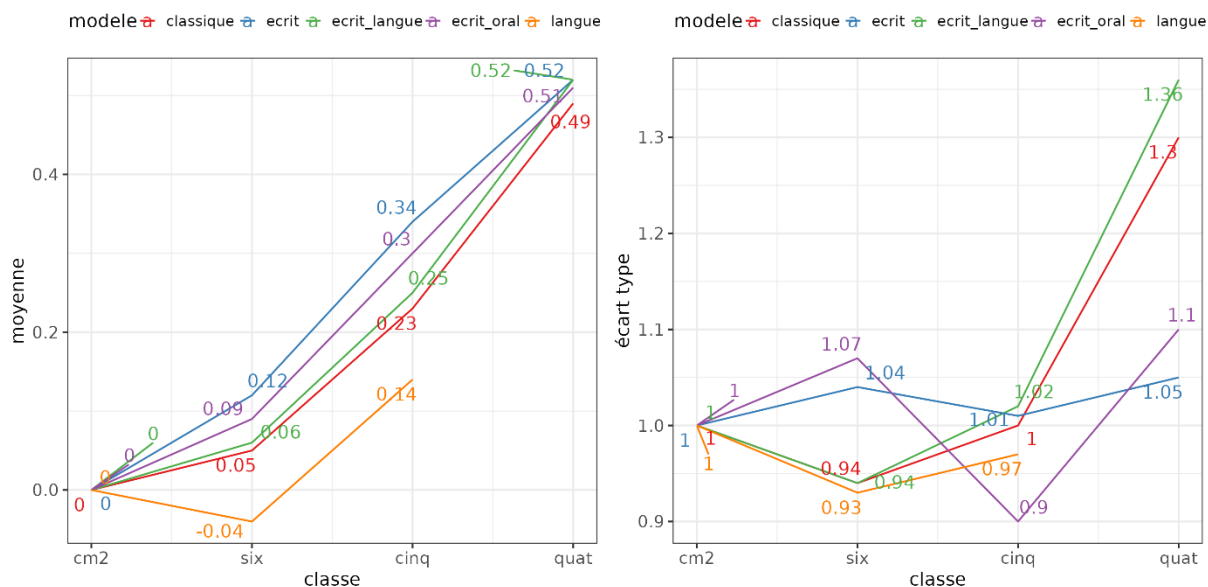
Enfin, les compétences évaluées en français en quatrième diffèrent sensiblement des autres niveaux puisqu'il n'y a pas d'exercices de grammaire ni d'orthographe dans ce test. Ces différences dans le contenu des épreuves ne perturbent pas l'estimation de la moyenne des scores, et donc l'évolution moyenne du niveau, mais conduisent à surestimer artificiellement l'évolution de la variance en quatrième. Pour y remédier, nous réduisons l'écart-type des scores de quatrième à sa valeur lorsque l'analyse est effectuée sur des évaluations au contenu strictement comparable du CM2 à la quatrième, c'est-à-dire aux exercices de compréhensions écrites et orales. Les tests de robustesse entrepris sur ce point et présentés ci-dessous (sous-partie B.5) confortent notre méthode.

5. Tests de robustesse

Plusieurs tests de robustesse sont réalisés pour s'assurer de la pertinence de nos choix méthodologiques. Nous présentons ici l'analyse sur la comparabilité des résultats de français en quatrième. Dans la figure 8, les moyennes et écarts-types des scores à chaque niveau sont présentés, en fonction des domaines évalués : en rouge, les élèves sont évalués sur l'ensemble du test, en bleu,

que sur l'écrit, en orange que sur le fonctionnement de la langue, en vert sur les deux (donc en excluant la compréhension de l'oral), et en violet sur l'écrit et l'oral. Dans chaque cas, les scores sont calculés en calibration concurrente sur les quatre niveaux, en gardant tous les items d'ancrage pour simplifier les analyses.

FIGURE 8 • Évolution au cours du temps des moyennes et des écarts-types des scores de français selon les domaines évalués



Lecture : si l'on évalue les élèves en français que sur les questions de compréhension de l'écrit et de fonctionnement de la langue, leur niveau moyen en cinquième vaut 0,25 et l'écart-type est de 1,02.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011.

Source : DEPP, panel 2011.

Nous observons que les évolutions des moyennes diffèrent peu selon les domaines évalués, sauf pour le fonctionnement de la langue où le niveau diminue en sixième, mais cela résulte du passage du papier au numérique entre le CM2 et la sixième, comme nous l'avons vu précédemment. Toutefois, nous observons que l'écart-type augmente considérablement en quatrième dans les spécifications où les élèves sont évalués en fonctionnement de la langue (les courbes rouge et verte). Or le test de quatrième ne comporte aucun item de fonctionnement de la langue. Compte tenu de cette différence de contenu entre la quatrième et les autres niveaux, et du fait que l'explosion de l'écart-type n'est pas observée lorsque nous évaluons les élèves que sur l'écrit et l'oral (courbes bleue et violette), nous en déduisons que l'augmentation de l'écart-type en quatrième reflète la différence de construit entre la quatrième et les autres tests, et non une augmentation réelle des écarts de performance. Dans notre analyse, nous choisissons de réduire l'écart-type de quatrième à sa valeur lorsque les élèves sont évalués en compréhension écrite et compréhension orale.

➤ II Progression des élèves au collège

À quel rythme les élèves progressent-ils au collège ? Quels types de parcours scolaires se dessinent ? Quels sont les écarts de compétence entre les élèves ? En utilisant les scores calculés pour tous les élèves à chaque niveau scolaire, ainsi que les données contextuelles sur les élèves et leur environnement familial, nous allons apporter des réponses à ces questions.

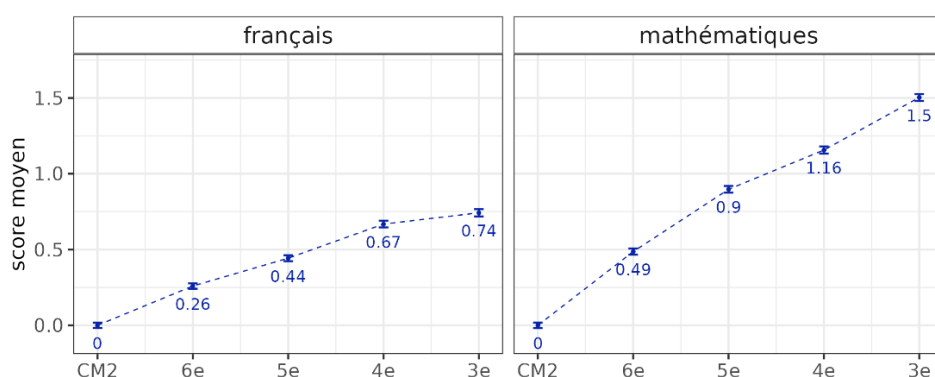
A. Progression moyenne et implications

1. Progression plus forte en mathématiques qu'en français

La figure 9 présente l'évolution des scores moyens des élèves au collège, avec comme référence les scores en fin de CM2 dont la moyenne a été fixée à 0 et l'écart-type à 1 dans chaque discipline. Les élèves progressent régulièrement chaque année, de manière plus importante en mathématiques qu'en français. En effet, la progression moyenne en français est d'un quart d'écart-type chaque année, alors qu'elle vaut plutôt un demi écart-type en mathématiques. La progression en mathématiques est dite plus rapide qu'en français, dans le sens où un élève ayant obtenu un score moyen de 0 en mathématiques en CM2 dépasse le niveau de CM2 des bons élèves (ceux ayant un score supérieur à 1, soit situés au-dessus du 85^e percentile) dès la quatrième, alors qu'en français, il n'atteint toujours pas ce niveau à la fin du collège. Par ailleurs, les résultats en mathématiques sont cohérents avec les progressions mesurées entre la sixième et la troisième avec les données du panel 2007, d'environ 0,9 écart-type (Ben Ali & Vourc'h, 2015).

Pour donner une illustration concrète de la signification de ces chiffres dans un contexte scolaire, prenons l'exemple d'une classe composée de 25 élèves, où ces derniers seraient classés selon leur score aux tests de CM2. Considérons un élève qui aurait la note médiane, c'est-à-dire qu'il aurait obtenu le 13^e meilleur score. S'il progressait comme la moyenne des élèves (i.e. de 0,26 écart-type en français et de 0,49 écart-type en mathématiques), il atteindrait en sixième le niveau en français qu'avait le 10^e élève en CM2, et en mathématiques le niveau qu'avait le 8^e élève en CM2³.

FIGURE 9 • Évolution du niveau des élèves au cours du collège



Lecture : le score moyen des élèves en français est de 0,44 en fin de cinquième, sachant que l'unité correspond à l'écart-type de la distribution des scores de CM2. Les petites barres verticales indiquent l'intervalle de confiance au seuil de 95 % sur la moyenne des scores des élèves. L'écart-type, réduit à 1 en CM2, demeure globalement stable à travers les niveaux.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois tests de compétence et complété l'un des questionnaires familles.

Source : DEPP, panel 2011.

³ Pour obtenir ce résultat, on considère que les scores des élèves suivent une loi normale. L'élève médian a le score moyen, l'élève classé 5^e a le score moyen plus un écart-type (car 34 % des élèves ont leurs notes comprises entre la moyenne et la moyenne plus un écart-type). De même, puisque 19 % des élèves ont une note entre la moyenne et la moyenne plus un demi écart-type, une progression de 0,49 écart-type pour le 13^e élève le conduit à la 8^e place environ.

2. Comparaison des progressions avec les États-Unis

La progression des élèves au collège peut être comparée avec celle des élèves aux États-Unis. En suivant la méthode utilisée par Bloom *et al.* (2008), nous calculons les gains de compétence des collégiens en France comparables aux indicateurs produits à partir de plusieurs tests standardisés proposés aux États-Unis.

Les résultats de notre étude sont très comparables avec ceux de Bloom *et al.* (2008), présentés dans le tableau 7. Les deux mêmes tendances se dégagent : les élèves ont des gains de scores plus importants en mathématiques qu'en langue, et ils progressent plus au début qu'à la fin du collège dans les deux disciplines. Toutefois, concernant le dernier point, la progression en français en classe de troisième est très nettement inférieure à celle des années précédentes, et également à son homologue étatsunienne. Cela pourrait venir du fait que le niveau de fin de troisième est mesuré par un test en début de seconde, soit après deux mois de vacances d'été qui peuvent entraîner une baisse des habitudes scolaires et donc des résultats. Mais comme cette relative baisse n'est pas observée en mathématiques, cette explication paraît peu satisfaisante.

TABEAU 7 • Progression moyenne en France et aux États-Unis

Niveau scolaire	Français / Reading		Mathématiques	
	France	États-Unis	France	États-Unis
CM2 - 6 ^e	0,27	0,32	0,47	0,41
6 ^e - 5 ^e	0,18	0,23	0,37	0,30
5 ^e - 4 ^e	0,21	0,26	0,23	0,32
4 ^e - 3 ^e	0,06	0,24	0,29	0,22

Lecture : l'augmentation du niveau est calculée entre des tests passés au printemps de chaque année scolaire. Les élèves en France progressent en moyenne de 0,47 écart-type en mathématiques en sixième, et de 0,41 aux États-Unis.

Afin d'établir des progressions comparables, les chiffres présentés pour la France sont obtenus en appliquant la formule de Bloom *et al.* (2008) :

$$r_{t-1,t} = \frac{\mu_t - \mu_{t-1}}{\sqrt{\frac{(n_{t-1}-1)\sigma_{t-1}^2 + (n_t-1)\sigma_t^2}{n_{t-1} + n_t - 2}}}$$

avec (μ_t, σ_t) la moyenne et l'écart-type de la distribution des scores au niveau scolaire t , et n_t le nombre d'élèves évalués cette année-là.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois tests de compétence et complété l'un des questionnaires familles, et élèves qui ont passé des tests standardisés aux États-Unis pendant les années 1990.

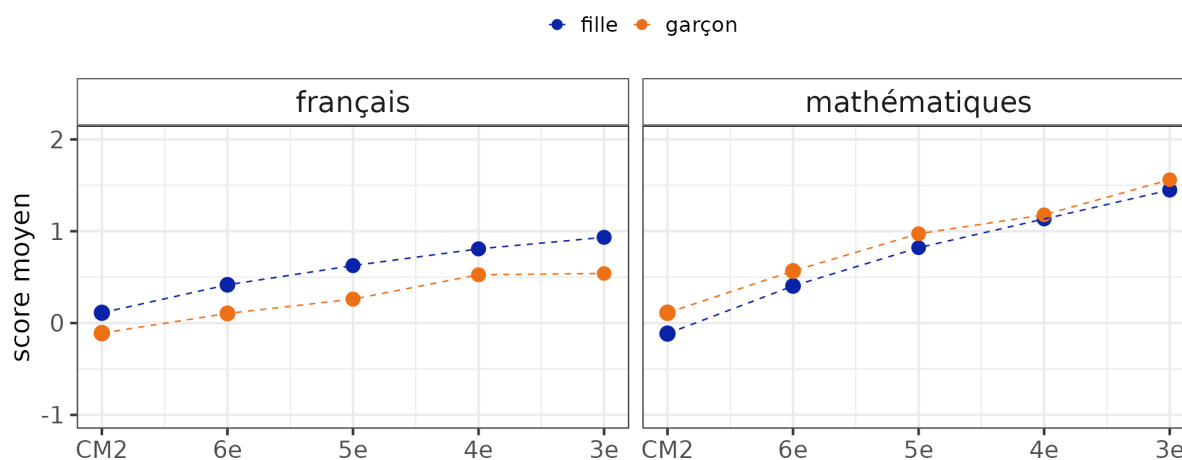
Source : DEPP, panel 2011 et Bloom *et al.* (2008).

La faible progression des élèves en troisième pourrait également être la conséquence de la crise sanitaire. En effet, les élèves du panel étaient en troisième pendant les premiers confinements de 2020 causés par l'épidémie de Covid-19, et donc leur scolarité a été fortement perturbée cette année-là. Plusieurs études internationales ont mesuré l'effet négatif des confinements sur les apprentissages des élèves, et certaines ont trouvé des effets plus forts en langue qu'en mathématiques (Maldonado & De Witte, 2022). Nos résultats en troisième suggèrent également que la progression des élèves a été ralentie par le confinement de 2020, surtout en français. En supposant qu'en l'absence de la crise sanitaire les élèves auraient progressé de la même manière qu'au cours des années précédentes, l'effet potentiellement attribuable au confinement serait de -0,16 écart-type en français et de -0,07 écart-type en mathématiques. Ces estimations sont très proches des effets du confinement du printemps 2020 évalués sur les apprentissages des élèves de CP en France, de -0,15 écart-type en français et -0,10 en mathématiques (Heidmann *et al.*, 2023).

B. Différences d'apprentissage entre les élèves

En plus de l'évolution globale des scores, il est possible d'observer comment les élèves progressent selon certaines caractéristiques. La figure 10 présente les résultats en distinguant les filles des garçons. Tout d'abord, les filles ont de meilleurs résultats en français dès le CM2 et inversement, en mathématiques, les garçons ont de meilleurs résultats dès le CM2. Toutefois, alors que l'écart semble légèrement se creuser en français au profit des filles, ces dernières rattrapent les garçons en mathématiques au cours du collège. Cette réduction des écarts au collège en mathématiques est notable car elle diffère du constat fait sur le primaire, où les écarts en mathématiques, inexistantes entre les filles et les garçons en début de CP, apparaissent dès le début de CE1 (Andreu *et al.*, 2023). Enfin, les écarts de compétence selon le sexe sont plus prononcés en français qu'en mathématiques, une observation déjà établie dans un état des lieux des disparités de compétence entre les sexes à travers un ensemble d'évaluations nationales et internationales (Chabanon, Steinmetz, 2018).

FIGURE 10 • Progression des filles et des garçons au collège



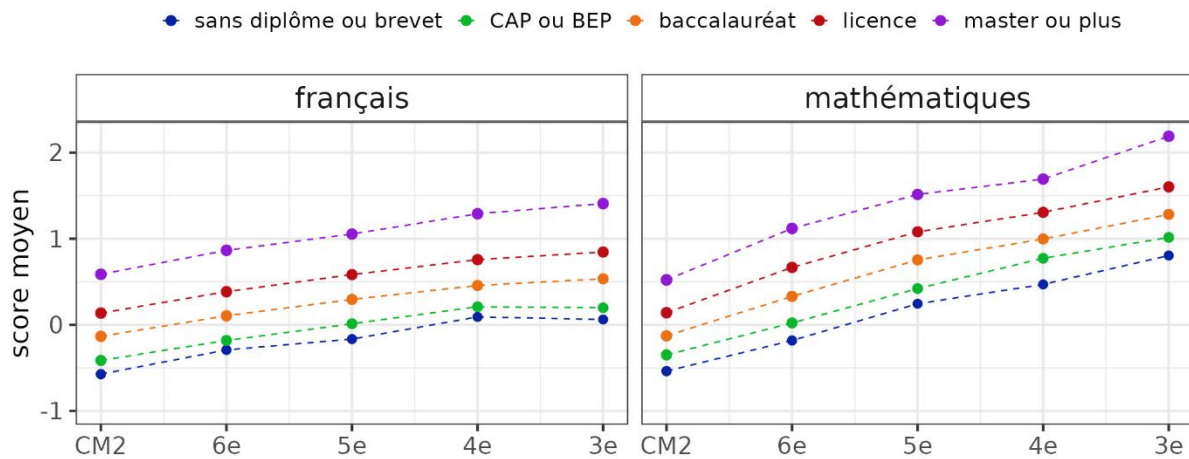
Lecture : les garçons ont un score moyen d'environ 1 en mathématiques en cinquième.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois tests de compétence et complété l'un des questionnaires familles.

Source : DEPP, panel 2011.

La figure 11 permet de visualiser les inégalités scolaires entre milieux sociaux en présentant l'évolution des scores en fonction du plus haut diplôme obtenu par les parents. Les écarts de performance sont frappants : en moyenne, les élèves dont les parents n'ont pas obtenu de baccalauréat n'atteignent jamais, au cours du collège, le niveau de français de fin de primaire des élèves dont les parents ont au moins un diplôme de master. Toutefois, un parallélisme des progressions moyennes entre les groupes d'élèves est globalement observé, à l'exception de la fin du collège. En français, les élèves les plus défavorisés connaissent un décrochage en classe de 3^e, tandis qu'en mathématiques, ce sont les élèves les plus favorisés qui enregistrent une progression plus marquée. Ces progressions plus différenciées en 3^e à l'avantage des élèves socialement favorisés pourraient être la conséquence de la crise sanitaire, les élèves du panel n'ayant jamais redoublé étant en 3^e pendant l'année scolaire 2019-2020. Une étude menée en Italie sur l'effet des fermetures d'écoles entre 2019 et 2021 sur des élèves en classe de 4^e (Carlana *et al.*, 2023) montre en effet que la pandémie de Covid-19 a non seulement entraîné une perte d'apprentissage global en français et en mathématiques mais a surtout creusé des écarts préexistants en fonction du statut socio-économique.

FIGURE 11 • Réussite scolaire des enfants selon celle des parents



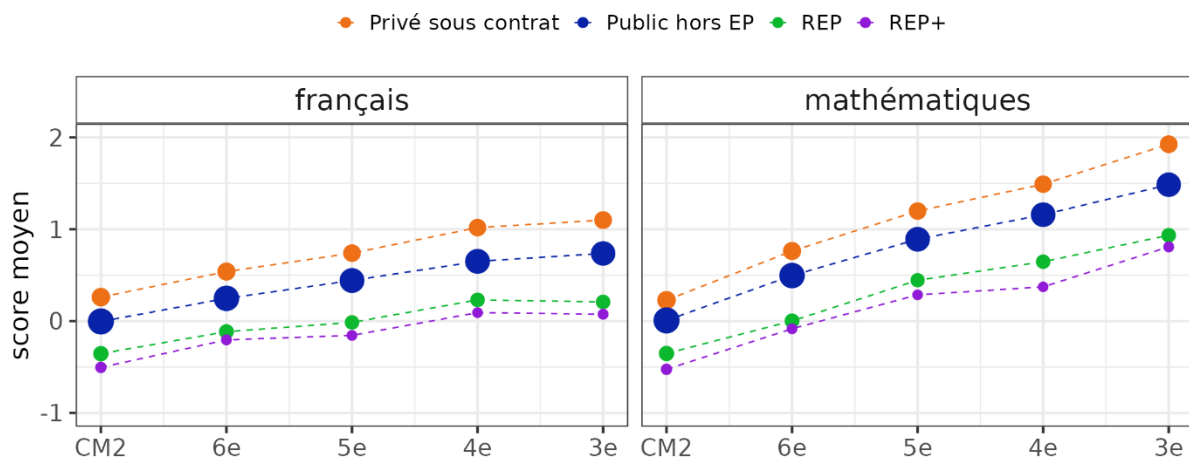
Lecture : le diplôme correspond au diplôme le plus élevé obtenu par les parents de l'élève. Les élèves dont les parents ont obtenu une licence ont un score moyen de mathématiques d'environ 1 en cinquième.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois tests de compétence et complété le premier questionnaire famille.

Source : DEPP, panel 2011.

Enfin, la figure 12 montre l'évolution des compétences des élèves selon le secteur de leur collège. Les élèves scolarisés dans les secteurs d'éducation prioritaire REP et REP+ arrivent au collège avec un moins bon niveau que leurs camarades du public hors éducation prioritaire, et ils progressent moins au collège, ce qui fait que les écarts de compétence se creusent. Les élèves scolarisés dans le secteur privé partent, quant à eux, avec de l'avance au début du collège et progressent plus vite en français et surtout en mathématiques.

FIGURE 12 • Évolution des résultats scolaires entre les secteurs



Lecture : dans la figure, la taille des points varie en fonction du nombre d'élèves scolarisés dans chaque secteur. Les élèves issus de collèges privés ont un score moyen de français d'environ 1 en quatrième.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois tests de compétence et complété l'un des questionnaires familles.

Source : DEPP, panel 2011.

Les écarts de performance entre secteurs sont donc très importants à la fin du collège : en fin de troisième, les élèves de REP+, en moyenne, n'atteignent pas en français le niveau de fin de CM2 des élèves du privé, et en mathématiques, leurs scores dépassent légèrement ceux de fin de sixième du privé. Cet écart plus important en français pourrait être dû à la crise sanitaire, car on a vu qu'elle avait surtout affecté les apprentissages de ce domaine et les élèves issus de milieux défavorisés sont les plus touchés, constat confirmé par plusieurs études internationales (Betthäuser *et al.*, 2023). Toutefois, le creusement des écarts est déjà visible en quatrième, avant la crise sanitaire. Il pourrait provenir des inégalités socio-économiques entre les différents secteurs, puisque les collèges privés

accueillent des élèves plus favorisés que dans le secteur public, et à l'inverse, l'éducation prioritaire est implantée dans les territoires défavorisés. Une analyse économétrique plus fine présentée dans la section suivante permettra de démêler l'effet du secteur de scolarisation de celui du milieu social sur les apprentissages des élèves.

➤ III Les facteurs de progression au collège

L'étude de l'évolution des scores a montré que les élèves ne progressent pas tous au même rythme au collège. Différents facteurs, comme le sexe ou le secteur de scolarisation, jouent un rôle dans leurs apprentissages. Afin d'estimer précisément les effets de chaque facteur, nous utilisons une approche économétrique basée sur les modèles de croissance. Ces modèles, cas particulier des modèles multiniveaux, ont l'avantage de modéliser la structure hiérarchique des données de panel, où plusieurs observations sont effectuées pour les mêmes unités à différents instants (Bressoux, 2010). En effet, les données sont structurées en deux niveaux : les scores y_{ti} obtenus par l'élève i au test du niveau scolaire t (niveau 1), et les élèves i (niveau 2).

ENCADRÉ Méthode d'estimation par modèles de croissance

Pour mesurer l'effet de chaque facteur sur la progression des élèves, plusieurs modèles ont été évalués. Le modèle 1 inclut le temps, c'est-à-dire la classe, comme seule variable explicative. Ce modèle s'écrit formellement au niveau 1 :

$$y_{ti} = \beta_{0i} + \beta_{1i}\text{temps}_t + e_{ti}$$

Et au niveau 2 :

$$\begin{aligned}\beta_{0i} &= \gamma_{00} + u_{0i} \\ \beta_{1i} &= \gamma_{10}\end{aligned}$$

avec y_{ti} le score obtenu par l'élève i au test de l'année t , temps_t une variable allant de 0 à 4 indiquant le niveau scolaire, u_{0i} l'erreur aléatoire associée à l'élève i supposée suivre une loi normale centrée et e_{ti} l'erreur aléatoire associée à chaque observation (à un temps t pour un élève i) supposée suivre également une loi normale centrée. Dans ce modèle, la progression moyenne des élèves au cours du collège est estimée par le coefficient γ_{10} . Plus précisément, la progression de l'élève i au cours du collège est modélisée par une droite de pente γ_{10} , commune entre tous les élèves, mais d'ordonnée à l'origine individualisée $\gamma_{00} + u_{0i}$, ce qui permet alors de modéliser les différences de compétence entre les élèves à l'entrée au collège. Les droites de régression estimées par ce modèle sont représentées dans le cadre de gauche de la figure 13.

Nous souhaitons ensuite rendre compte des différences de progression entre les élèves au collège en ajoutant des pentes aléatoires u_{1i} estimées par la régression suivante, au niveau 1 :

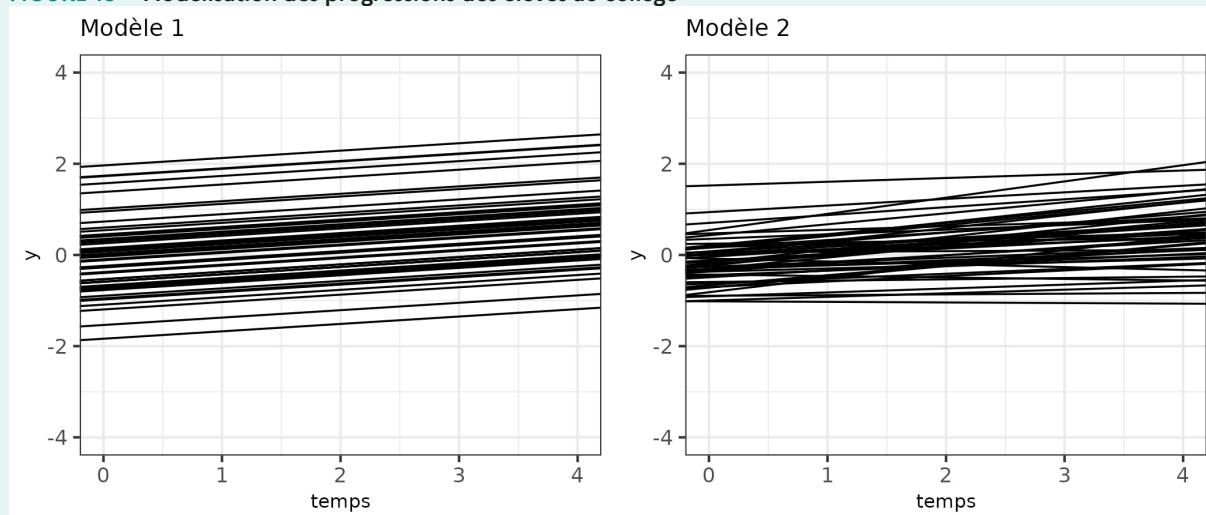
$$y_{ti} = \beta_{0i} + \beta_{1i}\text{temps}_t + e_{ti}$$

Au niveau 2 :

$$\begin{aligned}\beta_{0i} &= \gamma_{00} + u_{0i} \\ \beta_{1i} &= \gamma_{10} + u_{1i}\end{aligned}$$

Ainsi, l'évolution des scores de l'élève i au collège est modélisée par une droite de pente $\gamma_{10} + u_{1i}$ et d'ordonnée à l'origine $\gamma_{00} + u_{0i}$. Si le coefficient u_{1i} est positif, l'élève i progresse plus vite que la moyenne de ses camarades. Le cadre de droite de la figure 13 permet de visualiser les droites de régression estimées dans cette modélisation.

FIGURE 13 • Modélisation des progressions des élèves au collège



Lecture : la progression de chaque élève au cours du collège est modélisée par une droite, à constante aléatoire dans le modèle 1, et à pente et constante aléatoires dans le modèle 2. Les figures présentent ces droites pour 50 élèves pris au hasard.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois tests de compétence et complété l'un des questionnaires familles.

Source : DEPP, panel 2011.

Le modèle 2 inclut également plusieurs variables explicatives (sexe, redoublant et secteur), en interaction avec le temps afin de mesurer comment ces facteurs sont liés aux apprentissages des élèves. La régression s'écrit au niveau 1 :

$$y_{it} = \beta_{0i} + \beta_{1i} \text{temps}_t + e_{it}$$

Au niveau 2 :

$$\begin{aligned} \beta_{0i} &= \gamma_{00} + \gamma_{01} \text{Sexe}_i + \gamma_{02} \text{Redoublant}_{it} + \gamma_{03} \text{Secteur}_i + u_{0i} \\ \beta_{1i} &= \gamma_{10} + \gamma_{11} \text{Sexe}_i + \gamma_{12} \text{Redoublant}_{it} + \gamma_{13} \text{Secteur}_i + u_{1i} \end{aligned}$$

Dans ce modèle, $\gamma_{01}, \gamma_{02}, \gamma_{03}$ représentent les écarts initiaux de niveau entre élèves selon les variables explicatives, tandis que $\gamma_{11}, \gamma_{12}, \gamma_{13}$ estiment la progression supplémentaire des élèves selon ces variables.

La variable de redoublement a été incluse dans les modélisations en tant que variable de contrôle, en raison du design de l'évaluation où les élèves redoublants passent le test un an après le reste de l'échantillon. Cette variable indique si l'élève a un âge supérieur à l'attendu au moment de la passation de l'évaluation. Ainsi, la décision de redoublement est antérieure à l'année de mesure, elle n'est donc pas déterminée par le score de l'élève de l'année en cours. Toutefois, le coefficient associé à la variable ne mesure pas l'effet causal de redoubler une année, puisque les élèves redoublants diffèrent des autres élèves sur plusieurs caractéristiques, notamment sur leur niveau scolaire au moment du redoublement. Il faut considérer le caractère endogène dans les décisions de redoublement, lié au fait que les enseignants sélectionnent les élèves redoublants en fonction de leurs difficultés scolaires (Gary-Bobo et Robin, 2014).

Le modèle 3 est identique au modèle 2, mais comporte plus de variables explicatives permettant de décrire le milieu familial des élèves, grâce aux trois indices construits à partir des questionnaires, et prend en compte le niveau des élèves en CP.

Les hypothèses nécessaires à la bonne interprétation des coefficients de ces modèles ont été vérifiées et sont exposées après la présentation des résultats, dans la partie III.C.

A. Résultats empiriques

Les tableaux 8 et 9 présentent les résultats des régressions lancées séparément sur les scores de français et sur ceux de mathématiques. Chaque colonne correspond à une spécification différente, telles qu'elles sont énoncées ci-dessus.

La progression moyenne des élèves au collège est de 0,16 écart-type en français et de 0,34 en mathématiques, ce qui confirme les premières observations effectuées dans la partie précédente : les élèves progressent plus en mathématiques qu'en français au collège.

Le coefficient fortement négatif associé au redoublement (-0,42 en français et -0,49 en mathématiques, modèle 2) peut s'expliquer par le fait que ces élèves ont redoublé en raison de leurs

faibles performances initiales. Nous remarquons ainsi qu'une fois contrôlé du niveau en CP et de l'indice socio-économique (modèle 3), l'effet est considérablement atténué, mais demeure significatif.

TABLEAU 8 • Effets estimés par le modèle de croissance en français

	Modèle 1	Modèle 2	Modèle 3
Constante (γ_{00})	0,06**	- 0,03*	- 0,05**
Temps (γ_{10})	0,16**	0,14**	0,14**
Fille		0,23**	0,12**
Redoublant		- 0,42**	- 0,11**
Secteur privé		0,24**	0,04*
REP		- 0,34**	- 0,07**
REP+		- 0,51**	- 0,01
Score CP (français)			0,14**
Indice capital socio-éco			0,20**
Indice stratégies parents			0,08**
Indice soutien parents			- 0,08**
Temps*Fille		0,04**	0,04**
Temps*Secteur privé		0,04**	0,03**
Temps*REP		- 0,04**	- 0,02**
Temps*REP+		- 0,04**	- 0,01
Temps*Score CP			- 0,002*
Temps*Indice capital socio-éco			0,03**
Temps*Indice stratégies parents			0,01*
Temps*Indice soutien parents			- 0,003
Observations	54,739	54,739	54,739

Significativité : * au seuil de 10 % ; ** au seuil de 5 % ; *** au seuil de 1 %.

Lecture : selon le modèle 1, un élève progresse en moyenne de 0,16 écart-type chaque année. D'après le modèle 2, *ceteris paribus*, les filles progressent en moyenne de 0,04 écart-type en plus que les garçons chaque année.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois tests de compétence et complété l'un des questionnaires familles.

Source : DEPP, panel 2011.

TABLEAU 9 • Effets estimés par le modèle de croissance en mathématiques

	Modèle 1	Modèle 2	Modèle 3
Constante (γ_{00})	0,08**	0,24**	0,17**
Temps (γ_{10})	0,34**	0,32**	0,31**
Fille		- 0,24**	- 0,25**
Redoublant		- 0,49**	- 0,17**
Secteur privé		0,19**	0,03
REP		- 0,39**	- 0,18**
REP+		- 0,57**	- 0,18**
Score CP (mathématiques)			0,27**
Indice capital socio-éco			0,25**

Indice stratégies parents			0,06**
Indice soutien parents			- 0,10**
Temps*Fille		0,05**	0,05**
Temps*Secteur privé		0,06**	0,05**
Temps*REP		- 0,04**	- 0,02**
Temps*REP+		- 0,04**	- 0,01
Temps*Score CP			- 0,01**
Temps*Indice capital socio-éco			0,03**
Temps*Indice stratégies parents			0,01*
Temps*Indice soutien parents			- 0,002
Observations	52,848	52,848	52,848

Significativité : * au seuil de 10 % ; ** au seuil de 5 % ; *** au seuil de 1 %.

Lecture : toutes choses égales par ailleurs, selon le modèle 3, les élèves scolarisés en REP ont en moyenne, à l'entrée au collège, des scores inférieurs de 0,18 écart-type aux élèves du secteur public hors éducation prioritaire. Une augmentation d'un écart-type de l'indice de capital socio-économique entraîne, en moyenne, une hausse de 0,03 écart-type par année au collège.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois tests de compétence et complété l'un des questionnaires familles.

Source : DEPP, panel 2011.

1. Hétérogénéité des progressions selon le genre et le milieu scolaire au collège

Les filles arrivent au collège avec un meilleur niveau que les garçons en français (+ 0,23 écart-type, modèle 2), mais la situation est exactement inversée en mathématiques à l'avantage des garçons (- 0,24 écart-type). Fait intéressant, si on ajoute le niveau des élèves en CP (modèle 3), les filles perdent la moitié de leur avantage en français (+ 0,12) alors que les garçons conservent le leur en mathématiques (- 0,25). Cela suggère que l'avance des filles en français est en grande partie antérieure à l'arrivée en CP, alors que l'écart en mathématiques n'existe pas avant l'arrivée à l'école et se creuse au primaire. Ces résultats font écho à ceux constatés sur les données françaises de la cohorte Elfe, avant l'entrée à l'école élémentaire (Fischer et Thierry, 2022). Les auteurs observent l'absence d'écart en mathématiques en moyenne section de maternelle (et même un léger avantage pour les filles sur un échantillon non longitudinal) tandis qu'un avantage significatif en faveur des garçons apparaît au dernier trimestre de CP. En revanche, en compétences langagières, l'écart au profit des filles est déjà présent en maternelle. Ces différences dans le développement langagier entre filles et garçons avant même l'entrée à l'école ont également été observées dans de multiples études internationales (Rinaldi et al., 2021). Les explications fréquemment avancées ont trait à des différences biologiques et neuropsychologiques, amplifiées par les types d'interaction que les filles et les garçons entretiennent avec leurs parents, les activités proposées aux filles ayant un contenu langagier plus important. Quant aux observations faites à l'école primaire, elles rejoignent les résultats obtenus dans l'étude d'un panel d'écoliers suivis à partir de 1997 (Caille & Rosenwald, 2006). Les auteurs montrent que, durant cette période, les filles progressent davantage en français que les garçons et maintiennent donc leur avantage initial. En revanche, en mathématiques, les garçons parviennent à rattraper leur léger retard initial et même à creuser l'écart, en particulier vers la fin du primaire. Des tendances assez similaires ont été observées pour les élèves du panel 2011 en comparant leurs performances en CP et en CM2 (Fleury et al., 2022). Cette différenciation très précoce dans les apprentissages pourrait provenir de l'intériorisation des stéréotypes de genre qui associent les filles aux langues et les garçons aux sciences. Toutefois, notre étude montre que les écarts de performance entre filles et garçons évoluent différemment au collège. En effet, les filles progressent significativement plus que les garçons dans les deux disciplines (+ 0,04 écart-type en français et + 0,05 en mathématiques). Ainsi, contrairement à l'école primaire, le collège réduit les écarts entre filles et garçons en mathématiques. Ce résultat est en accord avec l'étude de Terrier (2020) qui, sur un échantillon de 4 500 élèves français, met en évidence des progrès supérieurs en mathématiques pour les filles au collège. Selon l'auteure, une partie de cet écart pourrait s'expliquer

par le fait que les enseignants de collège ont une attitude et des pratiques d'évaluation plus favorables aux filles.

Les élèves scolarisés dans le secteur privé ont de meilleurs résultats à l'entrée au collège que ceux du secteur public hors éducation prioritaire, aussi bien en français (+ 0,24, modèle 2) qu'en mathématiques (+ 0,19). Cependant, si on ajoute des informations plus précises sur le milieu socio-économique des élèves (modèle 3), l'écart entre les secteurs se réduit considérablement et devient même non significatif en mathématiques (+ 0,04 en français). Ainsi, les meilleures performances à l'entrée au collège des élèves du privé s'expliquent par leur environnement familial plus favorisé. Les différences sociales entre collèges publics et privés sont effectivement très fortes, et tendent même à s'accroître au cours des dernières années (Guillerm & Monso, 2022). Outre le critère socio-économique, les établissements privés effectuent une sélection à l'entrée basée sur la motivation et les résultats des élèves, à travers des entretiens avec les enfants et l'examen de leurs bulletins scolaires. Cette pratique peut être, elle aussi, source de disparités avec l'enseignement public à l'entrée du collège. Dès lors, le contrôle par le niveau en primaire et l'investissement des parents expliquerait également la baisse de l'effet du privé dans le modèle 3. Toutefois, le secteur privé a un effet positif sur la progression des élèves, même à contexte social, familial et scolaire égal entre élèves, ce qui veut dire que les établissements privés permettent aux élèves de progresser plus vite, surtout en mathématiques (+ 0,03 en français et + 0,05 en mathématiques). Une fois de plus, nos résultats vont à l'encontre de ce qui est observé dans l'enseignement primaire, où des études telles que celle de Fougère *et al.* (2017) ne montrent aucun effet favorable du secteur privé sur la progression des élèves. Ainsi, en accueillant des élèves plus favorisés et en leur offrant de meilleures conditions d'apprentissage, le secteur privé accroît les inégalités scolaires au collège. Ce résultat est en accord avec une étude réalisée sur un panel de collégiens lancé en 2007 (Moulin, 2022). À partir des résultats des élèves évalués en sixième et en troisième, l'auteur mesure des écarts d'acquis de l'ordre de 0,20 écart-type en fin de collège entre le privé et le public, en cohérence avec le différentiel de progression de 0,05 par année de collège que nous obtenons.

Le modèle 2 indique que les élèves scolarisés en éducation prioritaire progressent moins que les autres élèves. Toutefois, l'ajout de l'environnement social dans le modèle 3 permet d'expliquer pour moitié cette moindre progression pour les élèves de REP, et rend non significative les différences de progression entre collèges REP+ et collèges publics hors éducation prioritaire. Cela veut dire que les collèges de REP+ permettent à leurs élèves d'apprendre au même rythme que les autres élèves du public, sans réduire ni augmenter les inégalités scolaires. Un effet négatif demeure néanmoins pour les élèves de REP lorsqu'on inclut le milieu social (- 0,02 dans les deux disciplines). Les établissements de REP creusent les inégalités scolaires avec le public, possiblement à cause des effets de composition des collèges défavorisés qui peuvent induire des climats anti-scolaires, et des moyens dédiés insuffisants pour contrebalancer ces effets. Cependant, les résultats empiriques ne sont pas convergents sur le sujet et ne permettent pas de quantifier la part des effets de pairs, résultant des interactions entre élèves, dans l'effet négatif mesuré en REP (Monso *et al.*, 2019). Dans l'ensemble, ces constats sont concordants avec plusieurs études menées sur l'éducation prioritaire, qui suggèrent que les performances des élèves qui y sont scolarisés au collège sont peu différentes des autres élèves scolarisés dans le public, à caractéristiques sociales comparables (Caille *et al.*, 2016 ; Stéfanou, 2017).

2. Creusement des inégalités au collège

L'environnement socio-économique des élèves est le facteur le plus influent, parmi ceux retenus dans cette étude, sur les scores des élèves à l'entrée au collège en français (+ 0,20 pour un écart-type de plus dans l'indice socio-économique), et en mathématiques à égalité avec le niveau de CP (+ 0,25 et + 0,27). Les performances en mathématiques sont plus socialement déterminées que celles de français. En plus d'être meilleurs à l'entrée au collège, les élèves plus favorisés progressent plus vite que les autres (+ 0,03 dans les deux disciplines). Les inégalités de réussite scolaire continuent donc de se creuser au collège. Ce constat est concordant avec celui fait dans de nombreuses études (Duru-

Bellat & Mingat, 1988 ; Peugny, 2013 ; Ben Ali & Vourc'h, 2015 ; Cayouette-Remblière & Moulin, 2019). Toutefois, nos résultats montrent que l'essentiel des inégalités scolaires sont déjà établies avant l'entrée au collège, puisque l'effet cumulé à la fin du collège est de + 0,12 pour un écart-type de plus d'indice socio-économique (+ 0,03 chaque année, cumulé sur quatre ans), à comparer aux écarts initiaux de + 0,20 en français et + 0,25 en mathématiques.

Enfin, les élèves avec un score plus élevé en début de primaire arrivent au collège avec un niveau encore plus avancé (+ 0,14 en français et + 0,27 en mathématiques), synonyme d'un accroissement des écarts au primaire. Cependant, cette tendance s'inverse au collège. L'effet croisé du temps avec le score de CP est significativement négatif ce qui suggère qu'à toutes autres caractéristiques égales, les élèves initialement en difficulté tendent à légèrement rattraper leur retard au cours du collège.

Les deux indices sur les stratégies des parents et leurs implications personnelles ont des effets contraires sur les performances de leurs enfants. Les stratégies mises en œuvre par les parents sont corrélées avec de meilleurs résultats en fin de CM2 (+ 0,08 en français et + 0,06 en mathématiques), et associées à une amélioration légère tout au long du collège (+ 0,01 de progression dans les deux disciplines). À l'inverse, l'implication des parents est associée à de plus faibles performances en fin de primaire (- 0,08 et - 0,10), ce qui suggère que les parents s'investissent davantage lorsque leurs enfants rencontrent des difficultés à l'école. Dalsheimer-Van Der Tol et Murat ont montré, à partir des données du panel d'élèves entrés en sixième en 2007, que les élèves dont les parents rencontrent plus souvent les enseignants (*i.e.* au-delà des réunions parents-enseignants régulièrement organisées par le collège), sont aussi ceux qui ont le plus de difficultés à l'occasion des évaluations standardisées de fin de sixième. Or, la rencontre parents-enseignants est une variable incluse dans notre indice « soutien des parents ». Cette corrélation pourrait expliquer le coefficient négatif de l'indice, en raison d'une source potentielle d'endogénéité : les rencontres entre parents et enseignants sont suscitées par des difficultés scolaires. Toutefois, cette implication des parents en CM2 n'a pas d'effet à moyen terme sur la progression de leurs enfants au collège. Ce résultat rejoint d'autres études peu conclusives sur les effets de l'implication des parents sur les acquis cognitifs, même lorsqu'ils sont mesurés à court terme dans un cadre expérimental. En particulier, une étude a évalué les effets d'un programme d'information mis en place en France auprès de parents pour les aider à mieux accompagner leurs enfants dans leur scolarité (Avvisati *et al.*, 2013). Des effets positifs ont été observés sur l'implication effective des parents, le comportement des élèves, certaines notes données par les enseignants, mais pas sur les acquis cognitifs mesurés lors de tests standardisés comme dans notre étude. Il faut toutefois souligner que ces deux indicateurs sont construits à partir d'informations recueillies en fin de CM2, et traduisent une intention des parents qui n'a pas nécessairement perduré tout au long du collège, contrairement aux variables socio-économiques moins susceptibles de fluctuer au cours du temps. Cela pourrait expliquer, en partie, le faible impact de ces indicateurs sur les performances des élèves au collège.

B. Répartition de la variance

Les modèles de croissance mobilisés dans cette étude permettent de distinguer la variance des observations intra-élèves, reliée à l'évolution des scores dans le temps, de la variance interélèves, liée aux caractéristiques individuelles de chacun.

Les tableaux 10 et 11 présentent la décomposition de la variance entre ceux deux composantes, interélèves et intra-élèves, respectivement pour le français et les mathématiques.

Dans le modèle vide, on peut écrire :

$$\text{var}(Y_{it}) = \sigma_u^2 + \sigma_e^2$$

où σ_u^2 est la variance de l'erreur associée à chaque groupe (interélèves), et σ_e^2 est la variance de l'erreur associée à chaque observation (intra-élèves).

Entre le modèle vide et le modèle 1, la variance intra-élèves diminue, notamment en mathématiques où les progressions dans le temps sont plus fortes. En effet, la variance individuelle est expliquée par l'évolution dans le temps.

À partir du moment où l'on introduit des variables explicatives, la variance devient une fonction quadratique du temps :

$$\text{var}(Y_{ti}|\text{TEMPS}_t) = \sigma_e^2 + \sigma_{u0}^2 + \sigma_{u1}^2 \text{temps}_t^2 + 2\sigma_{u01} \text{temps}_t$$

où

- σ_e^2 décrit la variance associée à l'erreur intra-élèves ;
- σ_{u0}^2 est la variance de l'écart d'un individu à la constante moyenne dans la population ;
- σ_{u1}^2 est la variance de l'écart d'un individu à la pente moyenne de la population ;
- σ_{u01} est la covariance entre l'erreur sur les constantes et l'erreur sur les pentes.

TABLEAU 10 • Répartition de la variance en français

Type d'effets	Effets aléatoires	Modèle vide	Modèle 1	Modèle 2	Modèle 3
Interélèves	Variance des constantes	0,848	0,839	0,618	0,335
Interélèves	Variance des pentes			0,019	0,018
Interélèves	Covariance constantes/pentes			0,013	0,007
Intraélèves	Variance des constantes	0,367	0,302	0,253	0,253

Lecture : la variance des constantes interélèves pour le modèle vide en français vaut 0,848.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois tests de compétence et complété l'un des questionnaires familles.

Source : DEPP, panel 2011.

TABLEAU 11 • Répartition de la variance en mathématiques

Type d'effets	Effets aléatoires	Modèle vide	Modèle 1	Modèle 2	Modèle 3
Interélèves	Variance des constantes	0,793	0,817	0,649	0,416
Interélèves	Variance des pentes			0,016	0,015
Interélèves	Covariance constantes/pentes			0,005	0,0003
Intraélèves	Variance des constantes	0,691	0,392	0,349	0,349

Lecture : la variance des constantes intra-élèves pour le modèle vide en mathématiques vaut 0,691.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois tests de compétence et complété l'un des questionnaires familles.

Source : DEPP, panel 2011.

Lorsque l'on introduit les premières variables explicatives (sexe, redoublant et secteur) dans le modèle 2, la variance des constantes interélèves diminue fortement (de 26 % pour le français et de 21 % pour les mathématiques). Les écarts initiaux entre élèves sont fortement liés aux caractéristiques intrinsèques de chacun dont les covariables se font un facteur explicatif important. Dès lors, l'inclusion des indices sur l'environnement des élèves et leur score en CP dans le modèle 3 explique une part supplémentaire encore plus importante de la variance interélèves (baisse de 34 % supplémentaire pour le français et 29 % pour les mathématiques). Il est important de noter qu'il subsiste une part de variance, à la fois intra-élèves et interélèves, qui demeure inexpliquée. Ce constat est rassurant, suggérant que les différences de niveau entre les élèves ne peuvent pas être entièrement déterminées par leur niveau en CP et leurs caractéristiques socio-économiques, entre autres. Par exemple, les ressentis des élèves pourraient participer à réduire davantage ces variations initiales.

De plus, la variance des pentes demeure différente de zéro (0,018 en français et 0,015 en mathématiques). Cette dernière représente la part non expliquée de l'hétérogénéité dans les progressions, au-delà notamment des facteurs socio-économiques. Elle met en lumière la multiplicité

des rythmes d'apprentissage et souligne l'importance de tenir compte de cette variabilité dans une analyse qui pourrait être faite des différents profils de progression.

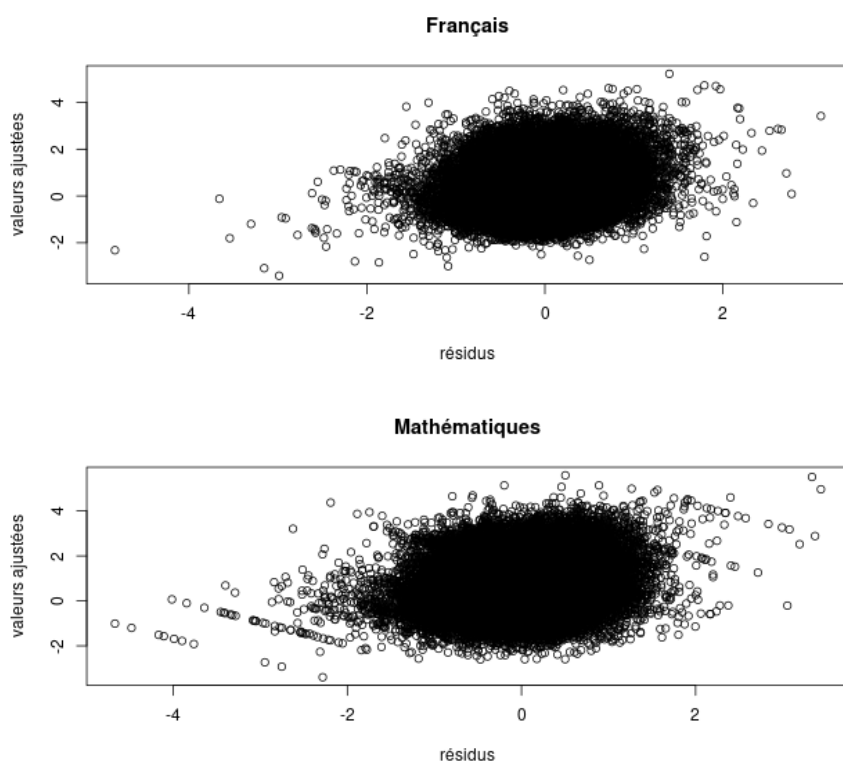
C. Hypothèses du modèle

Le modèle multiniveau tel que nous l'implémentons avec constantes et pentes aléatoires repose sur des hypothèses similaires à la régression linéaire multiple, en y incorporant cependant certaines modifications pour tenir compte de sa nature hiérarchique à deux niveaux et des effets aléatoires. Nous nous référons ici aux hypothèses telles qu'elles ont pu être définies par Raudenbush et Bryk (2002, p. 255).

1. Linéarité

L'hypothèse de linéarité implique que la moyenne de la variable réponse soit une combinaison linéaire des coefficients de régression et des variables prédictives. L'idée est de vérifier l'ajustement des données à un modèle linéaire, ce qui peut être perçu visuellement au moyen d'un nuage de points montrant les relations entre les résidus et les valeurs prédites par le modèle. Ce graphique ne doit présenter aucune tendance, *i.e.* les résidus doivent être dispersés de manière aléatoire autour de l'axe horizontal. C'est globalement ce qu'on observe dans la figure 14 : nous pouvons supposer être dans le cas d'une relation linéaire.

FIGURE 14 • Hypothèse de linéarité en français et en mathématiques



Lecture : pour chaque individu est représentée sa valeur prédite selon son résidu.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois tests de compétence et complété l'un des questionnaires familles.

Source : DEPP, panel 2011.

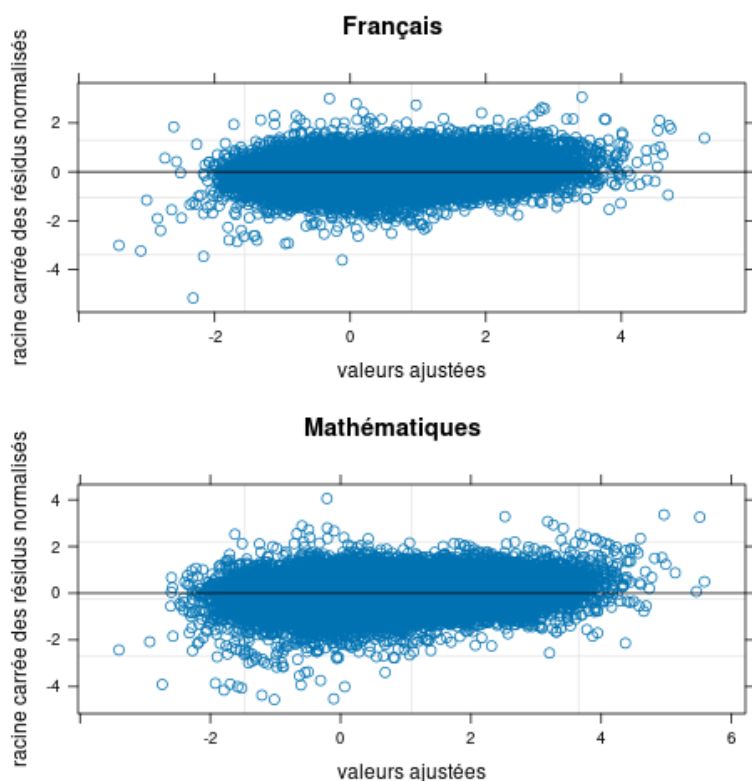
2. Homoscédasticité

L'hypothèse d'homogénéité des résidus consiste à vérifier que la variance des résidus est constante et ne dépend pas des valeurs des variables explicatives :

$$E(e_{it}^2) = \sigma_e^2$$

Elle peut être évaluée graphiquement en examinant la répartition de la racine carrée des résidus normalisés le long des valeurs prédites du modèle. Une ligne horizontale avec des points également répartis le long des prédicteurs est une bonne indication de l'homoscédasticité. C'est le cas dans la figure 15, où nous n'observons pas de signe d'hétéroscédasticité.

FIGURE 15 • Hypothèse d'homoscédasticité en français et en mathématiques



Lecture : pour chaque individu est représentée la racine carrée de son résidu normalisé selon sa valeur prédite.

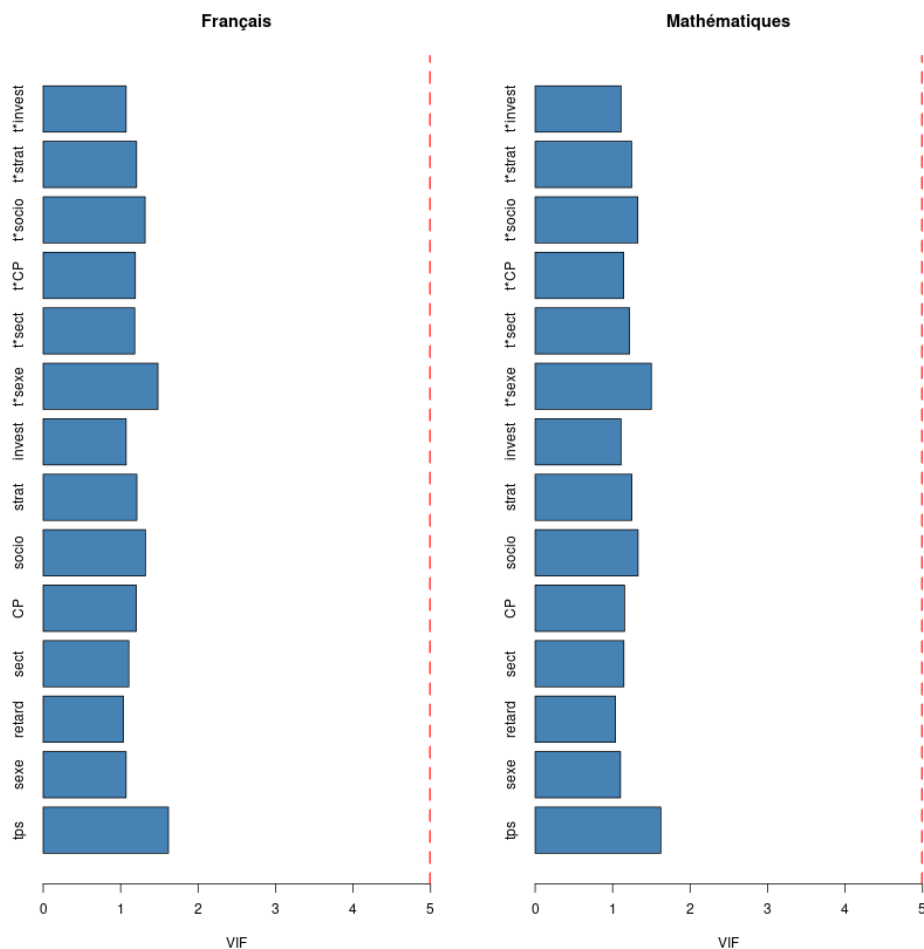
Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois tests de compétence et complété l'un des questionnaires familles.

Source : DEPP, panel 2011.

3. Absence de multicolinéarité

L'hypothèse de multicolinéarité stipule qu'aucune variable explicative incluse dans la régression ne peut être exprimée comme une combinaison linéaire des autres variables explicatives. Une évaluation de cette hypothèse peut être réalisée en calculant les facteurs d'inflation de la variance (VIF). Ces facteurs mesurent l'augmentation de la variance d'un coefficient en raison d'une relation linéaire avec d'autres variables explicatives. Dans le cas d'une absence de multicolinéarité, les VIF doivent être proches de 1, avec un seuil habituellement fixé à 5. La figure 16 présente les VIF pour chaque variable, et l'on peut observer qu'ils sont quasiment tous égaux à 1, ce qui conforte l'absence de multicolinéarité.

FIGURE 16 • Absence de multicolinéarité en français et en mathématiques



Lecture : la variable sexe a un VIF d'environ 1,1 dans notre modèle en mathématiques.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois tests de compétence et complété l'un des questionnaires familles.

Source : DEPP, panel 2011.

4. Hypothèses propres au modèle à deux niveaux

La structure hiérarchique des données entraîne l'existence de résidus distincts au niveau individuel (niveau 1) et au niveau du groupe (niveau 2 *i.e.* regroupés par élève), ce qui implique l'ajout d'hypothèses spécifiques pour chaque niveau. Dans notre contexte, les résidus au niveau 2 se situent au niveau des caractéristiques des élèves, indépendamment de leur évolution dans le temps. Par conséquent, chaque individu ne dispose plus de 3 à 5 points de mesure, représentant les différentes classes dans lesquelles il se trouve, mais d'un point unique. Au niveau 2, nous disposons de deux erreurs : u_{0i} , associée aux constantes, et u_{1i} , associée aux pentes.

Nous commençons par énoncer l'ensemble de ces hypothèses spécifiques :

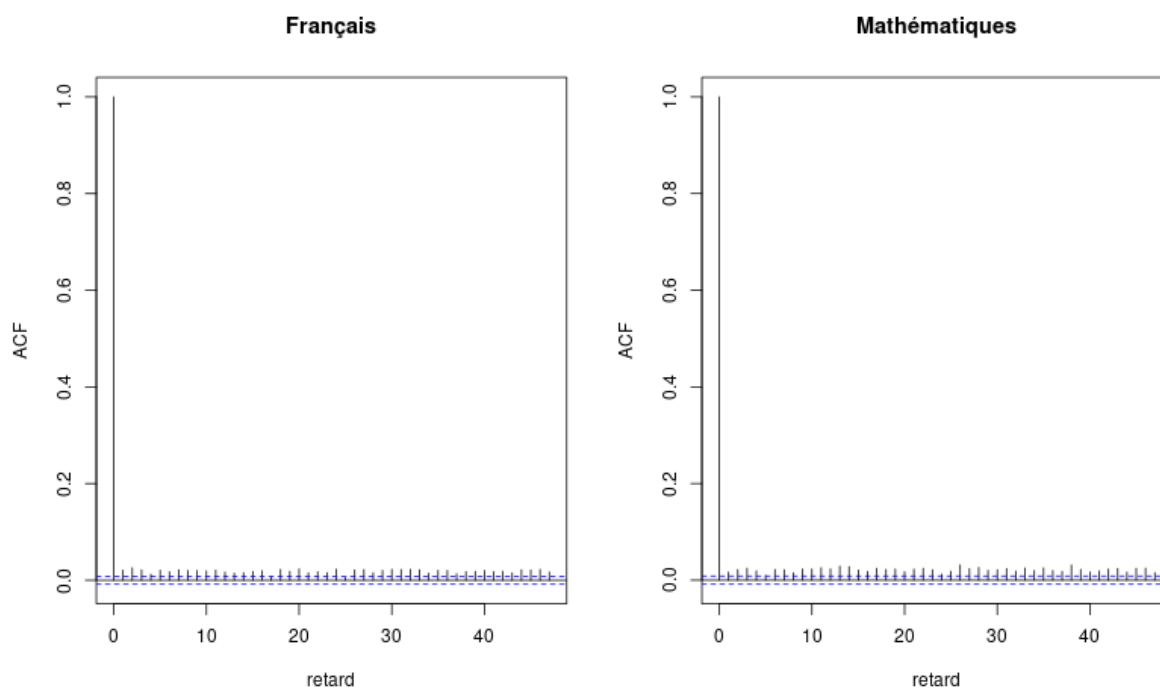
- Chaque erreur e_{it} de niveau 1 est supposée indépendante ($\mathbb{E}(e_t e_{t'}) = 0$ pour $t \neq t'$) et normalement distribuée de moyenne nulle et de variance σ_e^2 .
- Exogénéité : absence de corrélation entre les covariables et les caractéristiques individuelles non observées, mais qui ont un effet sur la variable d'intérêt (*i.e.* erreurs de niveau 1) : $\text{Cov}(X_i, e_{it}) = 0$.
- Les erreurs u_{0i} et u_{1i} de niveau 2 sont supposées indépendantes et de distribution normale multivariée avec une moyenne nulle et une variance de respectivement $\sigma_{u_0}^2$ et $\sigma_{u_1}^2$.
- Indépendance des variables explicatives de niveau 2 et des erreurs de niveau 2.
- Indépendance des erreurs entre les niveaux.

- Les variables explicatives de chaque niveau sont supposées non corrélées avec les termes aléatoires de l'autre niveau.

Considérant la taille importante de notre échantillon, la normalité des résidus de niveau 1 ou 2 peut être présumée en vertu du théorème de la limite centrale, qui assure la robustesse des tests d'hypothèses de Fisher et de Student (utilisés pour évaluer la significativité des coefficients) ainsi que pour la construction des intervalles de confiance, même en présence de légères déviations par rapport à la distribution normale. Un diagramme quantile-quantile et un histogramme de la distribution des résidus pourraient étayer cette assertion.

Étant donné la nature longitudinale (ou hiérarchique) des données, *i.e.* leur collecte à cinq reprises sur les mêmes individus au fil du temps (et donc organisées en groupes), il est possible qu'une dépendance apparaisse au niveau 1 entre les différentes mesures d'un même individu. Toutefois, le graphique des estimations de la fonction d'autocorrélation (ACF), présenté dans la figure 17, montre une absence d'autocorrélation des erreurs, justifiant du respect de l'hypothèse. Au niveau 2, les individus sont supposés indépendants entre eux⁴.

FIGURE 17 • Indépendance des erreurs de niveau 1



Lecture : la valeur en 0 représente l'autocorrélation de la série avec elle-même et vaut nécessairement 1. Pour les retards supérieurs à 0, les corrélations sont toutes très inférieures à 0,1, soit non significatives.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois tests de compétence et complété l'un des questionnaires familles.

Source : DEPP, panel 2011.

Ensuite, l'indépendance des variables explicatives et des erreurs aux niveaux 1 et 2 est supposée par construction du modèle de régression. De même, les erreurs e_{ti} de niveau 1 sont supposées indépendantes des erreurs u_{0i} et u_{1i} de niveau 2.

Enfin, notre dernier point implique l'absence de corrélation entre nos covariables de niveau individuel et les caractéristiques inobservées du groupe ayant un effet sur la variable d'intérêt. Cette hypothèse évalue notre choix d'un modèle à effets aléatoires par rapport à un modèle à effets fixes et peut être regardée pour les données observées au travers du test d'Hausman. Ce dernier compare l'estimation des coefficients sous le modèle à effets fixes et le modèle à effets aléatoires. Sous

⁴ Le tirage de l'échantillon a été effectué en sélectionnant des classes entières en CP. Toutefois, à l'entrée en CM2, les élèves sont globalement dispersés (4621 établissements différents).

l'hypothèse nulle H_0 , les deux estimateurs sont convergents mais l'estimateur sous le modèle à effets aléatoires est également efficace, donc préféré. Sous l'hypothèse alternative, seul l'estimateur du modèle à effets fixes converge. Ici, nous obtenons une statistique de test d'environ -550 et une p-valeur de 1. Nous ne rejetons donc pas l'hypothèse H_0 au seuil de 5 %. L'estimation par effets aléatoires est bien préférée.

D. Tests de robustesse

Nous avons réalisé des tests de robustesse en évaluant différentes spécifications pour garantir la fiabilité de notre modélisation principale. Les tableaux 12 et 13 présentent les résultats respectivement en français et en mathématiques.

Le modèle 1 correspond ici au modèle vide ne prenant en compte aucune variable explicative, répartissant ainsi la variance des données entre sa part interélèves et intra-élèves. Le modèle 2, qui intègre les variables explicatives, présente des constantes et des pentes aléatoires. Il est ainsi similaire à notre modélisation finale, à une différence près : il ne prend pas en compte l'effet croisé du temps avec nos variables explicatives. Cette particularité nous permet de visualiser, par comparaison, l'influence de chaque facteur de manière globale sur l'ensemble du collège. Le modèle 3 est identique à notre modèle final, exposé dans le corps du texte (avec la même numérotation). Sa reproduction dans ces tableaux vise à faciliter la comparaison visuelle avec les autres modèles.

Les trois modèles restants sont réalisés dans le but d'évaluer la robustesse de notre approche. Le modèle 4 est construit de la même manière que le modèle 3, à la différence près qu'il se limite exclusivement aux élèves ayant participé aux cinq évaluations. Cette restriction vise à maintenir un champ similaire pour tous les niveaux. Les coefficients obtenus sont très proches de ceux du modèle 3, bien que légèrement moins marqués, ce qui renforce la validité de notre approche. Ces faibles variations peuvent être attribuées à l'exclusion d'élèves susceptibles de présenter une progression moins prononcée : redoublants et élèves partis en apprentissage ou déscolarisés après la troisième.

De façon similaire, le modèle 5 restreint l'étendue des données, en excluant en plus les évaluations de troisième. Les statistiques descriptives ont mis en évidence un ralentissement des progressions dans les résultats en fin de troisième, que nous avons interprété comme la conséquence de la crise sanitaire. Par conséquent, ce modèle vise à évaluer les facteurs de progression moyenne au collège, en écartant les effets possibles du confinement du printemps 2020 sur les progressions des élèves. La progression moyenne au collège est effectivement plus significative, bien que l'écart ne soit pas considérable. Mais les coefficients des facteurs de progression sont très similaires, ce qui conforte une fois de plus notre modèle.

Enfin, pour le modèle 6, nous avons effectué une standardisation des scores à chaque niveau scolaire. Cette étape vise à confirmer la robustesse de nos résultats lorsque nous examinons les progressions relatives plutôt que les progressions absolues. Les progressions relatives fournissent moins de détails sur l'évolution des niveaux des élèves, mais elles ne dépendent pas des méthodes de comparaison des niveaux des élèves dans le temps. Ce test de robustesse nous permet donc de confirmer que les résultats obtenus ne sont pas influencés par nos choix méthodologiques dans les analyses psychométriques, mais qu'ils reflètent bien les tendances observées même avec des modèles plus simples. Encore une fois, les coefficients des facteurs de progression présentent une forte similitude. Une seule différence est notable : le coefficient du temps passe de +0,14 dans notre spécification choisie (modèle 3) à -0,04 en français, et de +0,31 à -0,05 en mathématiques. Cette différence peut s'expliquer par le fait que dans le modèle 6, tous les scores ont été centrés à chaque niveau scolaire, ce qui induit une progression nulle. De plus, la valeur est légèrement négative, puisque l'élève de

référence est un garçon et, comme observé précédemment, leur progression est moindre comparativement aux filles.

TABLEAU 12 • Coefficients estimés par le modèle de croissance en français

	Modèle 1	Modèle 2	Modèle 3	Modèle 4 (robustesse)	Modèle 5 (robustesse)	Modèle 6 (robustesse)
Constante	0,36**	- 0,08**	- 0,05**	0,04**	- 0,02	- 0,10**
Temps		0,16**	0,14**	0,13**	0,18**	- 0,04**
Fille		0,16**	0,12**	0,09**	0,12**	0,13**
Redoublant		- 0,12**	- 0,11**			- 0,09**
Secteur privé		0,06**	0,04*	0,04	0,04	0,04**
REP		- 0,10**	- 0,07**	- 0,11**	- 0,09**	- 0,08**
REP+		- 0,04	- 0,01	- 0,06	- 0,06	- 0,01
Score CP		0,14**	0,14**	0,14**	0,14**	0,15**
Indice capital socio-éco		0,23**	0,20**	0,18**	0,18**	0,21**
Indice stratégies parents		0,08**	0,08**	0,07**	0,07**	0,08**
Indice soutien parents		- 0,08**	- 0,08**	- 0,07**	- 0,07**	- 0,08**
Temps*Fille			0,04**	0,05**	0,02**	0,03**
Temps*Secteur privé			0,03**	0,02**	0,03**	0,02**
Temps*REP			- 0,02**	- 0,01	- 0,03*	- 0,02*
Temps*REP+			- 0,01	0,01	0,0003	- 0,01
Temps*Score CP			- 0,002*	- 0,001	- 0,01**	- 0,01**
Temps*Indice capital socio-éco			0,03**	0,03**	0,02**	0,01**
Temps*Indice stratégies parents			0,01*	0,01*	0,002	0,001
Temps*Indice soutien parents			- 0,003	- 0,005	- 0,002	0,002
Observations	54,739	54,739	54,739	29,935	23,948	54,739

Significativité : * au seuil de 10 % ; ** au seuil de 5 % ; *** au seuil de 1 %.

Lecture : toutes choses égales par ailleurs, selon le modèle 4, un élève progresse en moyenne de 13 % d'écart-type chaque année. Une fille a 9 % d'écart-type en plus, en moyenne, par rapport à un garçon.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois ou cinq tests de compétence et complété l'un des questionnaires familles.

Source : DEPP, panel 2011.

TABLEAU 13 • Coefficients estimés par le modèle de croissance en mathématiques

	Modèle 1	Modèle 2	Modèle 3	Modèle 4 (robustesse)	Modèle 5 (robustesse)	Modèle 6 (robustesse)
Constante	0,73**	0,13**	0,17**	0,33**	0,30**	0,09**
Temps		0,34**	0,31**	0,31**	0,34**	- 0,05**
Fille		- 0,18**	- 0,25**	- 0,31**	- 0,31**	- 0,24**
Redoublant		- 0,18**	- 0,17**	- 0,11*		- 0,18**
Secteur privé		0,09**	0,03	0,01	0,02	0,03
REP		- 0,21**	- 0,18**	- 0,19**	- 0,16**	- 0,17**
REP+		- 0,21**	- 0,18**	- 0,20**	- 0,15*	- 0,17**

Score CP		0,26**	0,27**	0,25**	0,26**	0,26**
Indice capital socio-éco		0,30**	0,25**	0,21**	0,22**	0,25**
Indice stratégies parents		0,07**	0,06**	0,08**	0,09**	0,06**
Indice soutien parents		- 0,10**	- 0,10**	- 0,10**	- 0,09**	- 0,09**
Temps*Fille			0,05**	0,05**	0,05**	0,05**
Temps*Secteur privé			0,05**	0,04**	0,03**	0,04**
Temps*REP			- 0,02**	0,0003	- 0,02	- 0,01
Temps*REP+			- 0,01	- 0,02	- 0,07**	- 0,01
Temps*Score CP			- 0,01**	- 0,01**	- 0,01**	- 0,01**
Temps*Indice capital socio-éco			0,03**	0,03**	0,01*	0,02**
Temps*Indice stratégies parents			0,01*	0,002	- 0,004	0,003
Temps*Indice soutien parents			- 0,002	- 0,002	- 0,003	0,001
Observations	52,848	52,848	52,848	27,79	22,232	52,848

Significativité : * au seuil de 10 % ; ** au seuil de 5 % ; *** au seuil de 1 %.

Lecture : toutes choses égales par ailleurs, selon le modèle 5, un élève scolarisé en REP en CM2 a en moyenne un score inférieur de 16 % d'écart-type à celui d'un élève dans le secteur public hors éducation prioritaire. Une augmentation d'un point de l'indice de capital socio-économique entraîne, en moyenne, une hausse de score de 1 % d'écart-type par année au collège.

Champ : panel d'élèves de France hors DROM entrés en CP en 2011 ayant réalisé au moins trois ou cinq tests de compétence et complété l'un des questionnaires familles.

Source : DEPP, panel 2011.

➤ Conclusion

Notre étude mesure les progressions des élèves au collège, année par année, en français et en mathématiques. Grâce aux tests standardisés spécialement conçus pour notre enquête, ainsi qu'à l'application de modèles de la théorie de réponse à l'item, nous avons comparé les compétences des élèves en CM2, avec leurs compétences en sixième, en cinquième, et ainsi de suite jusqu'à la classe de troisième. Cette approche nous a permis de calculer les gains de compétence entre deux niveaux scolaires. Sur l'ensemble du collège, nous avons une progression moyenne au cours d'une année scolaire de 0,34 écart-type en mathématiques et de 0,16 écart-type en français.

Les élèves n'apprennent pas tous au même rythme au collège. La richesse des données contextuelles recoltées lors de l'enquête a permis d'examiner cette hétérogénéité dans l'évolution des compétences. Nos résultats indiquent une progression plus forte des filles que des garçons en français et en mathématiques, d'environ + 0,05 écart-type par an, réduisant ainsi l'écart qui s'était creusé en mathématiques à l'école primaire, tandis que l'écart en français continue de s'accroître. Nous avons également montré que les élèves scolarisés dans un collège privé progressent plus que leurs camarades du public, toutes choses égales par ailleurs, surtout en mathématiques avec + 0,05 de progression par an. De plus, nos résultats ont mis en évidence que le collège ne parvient pas à réduire les inégalités de réussite entre les différents groupes sociaux. Les élèves issus de milieux défavorisés rencontrent davantage de difficultés dès leur entrée en sixième, et progressent moins vite que leurs camarades plus favorisés. Cette situation contribue ainsi à l'accentuation des inégalités scolaires au collège. Alors, dans la mesure où l'articulation entre ces différentes sources de disparités peut se faire cumulative, il serait intéressant d'examiner plus en détail des sous-ensembles d'élèves, tel que les filles dans le privé ou les élèves bénéficiant d'un soutien des parents en REP+, relativement au reste de cette population.

Aussi, la pluralité des parcours des élèves, que l'on a caractérisée par des pentes de progressions variables, pourrait dissimuler des trajectoires non linéaires chez certains groupes d'élèves. Une perspective de prolongement de cette étude consisterait à enrichir notre compréhension des trajectoires individuelles en les étudiant de manière continue sur l'ensemble du collège, autour d'une classification des différents types de progression. En regroupant les élèves en fonction de schémas de progression similaires, l'idée serait de pouvoir distinguer, par exemple, les élèves qui progressent régulièrement chaque année, de ceux qui progressent très vite au début et stagnent par la suite, et de ceux qui effectuent un bond de progression à la fin du collège, parmi tous les autres profils possibles.

Notre étude s'est heurtée à quelques difficultés qui méritent d'être rappelées. Tout d'abord, la crise sanitaire a perturbé l'enquête en annulant le dernier point de mesure prévu en troisième. Pour contourner ce problème, nous avons choisi d'utiliser les résultats des élèves du panel aux tests de positionnement de début de seconde, passés par tous les élèves entrant au lycée depuis 2018. Cependant, pour relier ces tests à ceux de notre enquête, nous avons dû recourir à un nouvel échantillon d'élèves entrés en seconde en 2022, donc deux ans plus jeunes que notre panel initial, ce qui pourrait induire un effet de génération non contrôlé dans nos analyses. De plus, l'utilisation d'un test en début de seconde pour évaluer le niveau de fin de collège présente deux limites : les élèves en apprentissage, en CAP, scolarisés hors Éducation nationale ou déscolarisés après la troisième ne sont pas évalués et, pour les élèves restants, les deux mois de vacances d'été ont pu affecter leurs performances scolaires. Néanmoins, les nombreux tests de robustesse que nous avons réalisés, ainsi que les comparaisons avec la littérature sur le sujet, confirment la cohérence de nos résultats et renforcent ainsi la validité de nos choix méthodologiques.

De plus, comme le panel d'élèves a été constitué au début de l'école primaire, nous ne disposons pas de classe entière au collège, les élèves étant éparpillés dans différentes classes ou différents établissements au fil des années, ce qui empêche d'explorer le rôle des camarades et de l'environnement scolaire sur la progression de chacun. En vue de prochaines études, la récente généralisation des évaluations nationales au collège présente une opportunité inédite pour étudier les effets de pairs. Grâce aux données à grande échelle, il serait en effet possible d'analyser l'influence des interactions entre pairs sur les trajectoires d'apprentissage des élèves et d'étudier comment s'articulent les contextes sociaux et scolaires pour favoriser ou entraver la réussite des élèves.

Enfin, l'étude de ce panel n'est pas terminée. Les élèves ont été suivis jusqu'en terminale et en appariant avec des données administratives, il sera possible de poursuivre jusqu'à leur insertion professionnelle. Notre mesure de la progression au collège sera alors un facteur important pour expliquer les différents parcours dans le supérieur et les choix de vie professionnelle.

RÉFÉRENCES DES PUBLICATIONS DE LA DEPP



Andreu, S., Conceicao, P., Desclaux, J., Eteve, Y., Laskowski, C., Le Breton, S., Neirac, L., Persem, E., Rocher, T., Rue, G., Thumerelle, J., Vourc'h, R., Wuillamier, P. (2023), "Évaluations Repères 2022 de début de CP et de CE1 : des résultats comparables à ceux de 2021, à l'exception d'une baisse en français en CE1", *Note d'Information*, 23.01, DEPP. <https://doi.org/10.48464/ni-23-01>

Andrieux, V., Chollet-Remvikos, P. Levasseur, J. (2002), "Avant et après les vacances, évolution des acquis des élèves", *Note d'Information*, 02.34, DEPP.

Avvisati F., Givord P., 2021, How much do 15-year-olds learn over one year of schooling? An international comparison based on PISA, *Documents de travail de l'OCDE sur l'éducation*, n° 257, Éditions OCDE, Paris, <https://doi.org/10.1787/a28ed097-en>.

Avvisati F., Givord P., 2023, The Learning Gain over One School Year among 15-year-olds: An International Comparison based on PISA. *Labour Economics*, 84, p.102365. <https://doi.org/10.1016/j.labeco.2023.102365>.

Avvisati F., Gurgand M., Guyon N., Maurin E., 2014, Getting Parents Involved: A Field Experiment in Deprived Schools, *The Review of Economic Studies*, 81(1), 57-83. <https://doi.org/10.1093/restud/rdt027>

Béland S., Cousineau D. Loye N., 2017, Utiliser le coefficient omega de McDonald à la place de l'alpha de Cronbach, *Revue des sciences de l'éducation de McGill*, 52(3), 791-804. <https://doi.org/10.7202/1050915ar>

Ben Ali L., Vourc'h R., 2015, Évolution des acquis cognitifs au collège au regard de l'environnement de l'élève : constat et mise en perspective longitudinale, *Éducation & formations*, 86-87, 211-233. <https://doi.org/10.48464/ef-86-87-10>

Bernigole V., Fernandez A., Loi M., Salles F., 2023, PISA 2022 : la France ne fait pas exception à la baisse généralisée des performances en culture mathématiques dans l'OCDE, *Note d'Information*, 23.48, DEPP. <https://doi.org/10.48464/ni-23-48>

Betthäuser B. A., Bach-Mortensen A. M., Engzell P., 2023, A systematic review and meta-analysis of the evidence on learning during the Covid-19 pandemic, *Nature Human Behaviour*, 7, 375-385. <https://doi.org/10.1038/s41562-022-01506-4>

Birnbaum A., 1968, Some latent trait models and their use in inferring an examinee's ability, in F. Lord, M. R. Novick., *Statistical theories of mental test scores* (p. 397-479), Addison-Wesley.

Bloom H., Hill C., Black A. R., Lipsey M. W., 2008, Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions, *Journal of Research on Educational Effectiveness*, 1(4), 289-328. <https://doi.org/10.1080/19345740802400072>

Bock R. D., Aitkin M., 1981, Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm, *Psychometrika*, 46(4), 443-459. <https://doi.org/10.1007/BF02293801>

Bourdieu P., Passeron J.-C., 1964, *Les Héritiers - Les étudiants et la culture*, Éditions de Minuit.

Bressoux P., 2010, *Modélisation statistique appliquée aux sciences sociales*, De Boeck supérieur. <https://doi.org/10.3917/dbu.bress.2010.01>

Bret A., Durand de Monestrol H., Hick M., Salles F., Fernandez A, Loi M., 2023, « PISA 2022 : culture scientifique, compréhension de l'écrit et vie de l'élève », *Note d'Information*, 23.49, DEPP. <https://doi.org/10.48464/ni-23-49>

Caille J.-P., 2017, Quarante-cinq ans de panels d'élèves à la Direction de l'évaluation, de la prospective et de la performance (DEPP), *Éducation & formations*, 95, pp.5-31. 10.48464/halshs-01760754

Caille J.-P., Davezies L., Garrouste M., 2016, Les résultats scolaires des collégiens bénéficient-ils des réseaux ambition réussite ? Une analyse par régression sur discontinuité, *Revue économique*, 67(3), 639-666. <https://doi.org/10.3917/reco.673.0639>

Caille J.-P., Rosenwald F., 2006, Les inégalités de réussite à l'école élémentaire : construction et évolution, *Insee références*.

Carlana M., La Ferrara E. & Lopez C., 2023, Exacerbated Inequalities: The Learning Loss from Covid-19 in Italy. *AEA Papers and Proceedings*. <https://doi.org/10.1257/pandp.20231010>

Cayouette-Remblière J., Moulin, L., 2019, Comment évoluent les inégalités de performances scolaires au collège ? Un suivi longitudinal des élèves entre la 6^e et la 3^e, *Population*, 74, 551-586. <https://doi.org/10.3917/popu.1904.0551>

Chabanon L., Jouvenceau M., 2022, De l'école élémentaire à l'entrée dans l'enseignement supérieur : filles et garçons construisent des parcours distincts, *Insee références*.

Chabanon L., Steinmetz C., 2018, Écarts de performances des élèves selon le sexe : que nous apprennent les évaluations de la DEPP ? *Éducation & formations*, 96, 39-57. <https://doi.org/10.48464/halshs-01824495>

Cibois P., 2014, Les figures de l'analyse des correspondances, in P. Cibois, *Les méthodes d'analyse d'enquêtes*, Lyon : ENS Éditions. <https://doi.org/10.4000/books.enseditions.1462>

Dalsheimer-Van Der Tol N., Murat F., 2011, Les parents et l'école en France et en Europe, *Éducation & formations*, 80, 79-94.

Delarue M.-C., Heidmann L., Raffy G., 2024, Apprentissages hétérogènes : comment les élèves progressent au collège ? Une étude psychométrique de l'évolution des compétences des élèves, *Éducation & formations*, 107, 7-34. <https://doi.org/10.48464/ef-107-061>

DeMars C. E., 2018, Classical Test Theory and Item Response Theory, in P. Irwing, T. Booth, D. J. Hughes, *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development* (p. 49-73), Wiley Blackwell. <http://dx.doi.org/10.1002/9781118489772.ch2>

Deroyon T., 2017, La correction de la non-reponse par repondération, Insee, département des méthodes statistiques.

Dos Santos R., Ninnin L., Paillet V., Salles F., 2023, Étude de comparabilité de passations « papier crayon » et numérique : l'exemple de Cedre mathématiques collège, *Éducation & formations*, 105, 117-142. <https://doi.org/10.48464/ef-105-06>

Duru-Bellat M., Mingat A., 1988, De l'orientation en fin de cinquième au fonctionnement du collège. Vol. 2. Progression, notation, orientation : l'impact du contexte de scolarisation, *Les cahiers de l'Irédu*.

Duru-Bellat M., Mingat A., 1993, *Pour une approche analytique du fonctionnement du système éducatif*, Paris, PUF.

Fischer J-P & Thierry X., 2022, Boy's math performance, compared to girls', jumps at age 6 (in the ELFE's data at least). *British Journal of Developmental Psychology*, 40, 504–519. <https://doi.org/10.1111/bjdp.12423>

Fleury D., Le Cam M., Vourc'h R., 2022, Panel des élèves entrés en CP en 2011 - Performances à l'école élémentaire selon le niveau scolaire initial et l'origine sociale, *Note d'Information*, 22.14, DEPP. <https://doi.org/10.48464/ni-22-14>

Fougère D., Monso O., Rain A., Tô M., 2017, Qui choisit l'école privée, et pour quels résultats scolaires ? *Éducation & formations*, 95, 59-85. <https://dx.doi.org/10.48464/halshs-01760816>

Gary-Bobo R. & Robin J., 2014, La question des redoublements : analyse économique et problèmes statistiques, *Revue économique*, 65, 5-45. <https://doi.org/10.3917/reco.651.0005>

Grisay A., 1997, *Évolution des acquis cognitifs et socio-affectifs des élèves au cours des années de collège*, dossier *Éducation & formations*.

Guillerm M., Monso O., 2022, « Évolution de la mixité sociale des collèges », *Note d'Information*, 22.26, DEPP. <https://doi.org/10.48464/ni-22-26>

Haziza D., Beaumont J.-F., 2007, On the construction of imputation classes in surveys, *International Statistical Review*, 75, 25-43.

Heidmann L., Neirac L., Andreu S., Conceicao P., Eteve Y., Fabre M., Vourc'h R., 2023, Delayed learning to read and write during the Covid-19 pandemic: longitudinal study of the heterogeneous effects on all first graders in France, *SocArXiv qn9a8*, Center for Open Science. <https://ideas.repec.org/p/osf/socarx/qn9a8.html>

Ichou M., 2013, Différences d'origine et origine des différences : les résultats scolaires des enfants d'émigrés/immigrés en France du début de l'école primaire à la fin du collège, *Revue française de sociologie*, 54, 5-52. <https://doi.org/10.3917/rfs.541.0005>

Lee W.-C., Lee G., 2018, *IRT linking and equating*. In P. Irwing, T. Booth, D. J. Hughes, *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development* (p. 639-673), Wiley Blackwell. <https://doi.org/10.1002/9781118489772.ch21>

Maldonado J. E., De Witte K., 2022, The effect of school closures on standardised student test outcomes, *British Educational Research Journal*, 48(1), 49-94. <https://doi.org/10.1002/berj.3754>

Monso O., Fougère D., Givord P., Pirus C., 2019, Les camarades influencent-ils la réussite et le parcours des élèves ? *Éducation & formations*, 100, 23-52. <https://dx.doi.org/10.48464/ef-100-02>

Moulin L., 2022, Do private schools increase academic achievement? Evidence from France, *Education Economics*, 31, 247-274. <https://doi.org/10.1080/09645292.2022.2061428>

OECD, 2020. « Construction of indices », in *PISA 2018 Results (Volume III): What School Life Means for Students' Lives*, OECD Publishing, Paris. <https://doi.org/10.1787/0a428b07-en>

Peugny C., 2013, *Le Destin au berceau. Inégalités et reproduction sociale*, Seuil, coll. « La République des idées ».

Rasch G., 1980, *Probabilistic model for some intelligence and achievement tests*, University of Chicago Press.

Raudenbush S. W., Bryk A. S., 2002, *Hierarchical linear models. Applications and data analysis methods* (2^e édition), Sage Publications.

Rinaldi P., Pasqualetti P., Volterra V. & Caselli M. C., 2023, Gender differences in early stages of language development. Some evidence and possible explanations, *Journal of Neuroscience Research*, 101, 643–653. <https://doi.org/10.1002/jnr.24914>

Rocher T., 2016, *Construction d'un indice de position sociale des élèves*, *Éducation & formations*, 90, DEPP, pp. 5-27. <https://doi.org/10.48464/hal-01350095>

Rocher T., 2013, *Mesure des compétences : les méthodes se valent-elles ? : questions de psychométrie dans le cadre de l'évaluation de la compréhension de l'écrit*, [Thèse de doctorat en Psychologie, université Paris Ouest Nanterre La Défense]. thèses.fr. <https://www.theses.fr/2013PA100176>

Stéfanou A., 2017, *Éducation prioritaire. Scolarité des élèves de 2007 à 2012*, *Éducation & formations*, 95, 87-106. <https://doi.org/10.48464/halshs-01779982>

Terrier C., 2020, Boys lag behind: How teachers' gender biases student achievement, *Economics of Education Review*, 77. <https://doi.org/10.1016/j.econedurev.2020.101981>

Trancart D., 2012, Quel impact des ségrégations socio-spatiales sur la réussite scolaire au collège ? *Formation emploi*, 120, 35-55. <https://doi.org/10.4000/formationemploi.3761>

Warm T.A., 1989, Weighted likelihood estimation of ability in item response theory, *Psychometrika*, 54, 427-450. <https://doi.org/10.1007/BF02294627>

Yin L., Fishbein B., Bezirhan U., Foy P. & von Davier M., 2023, Examining country-level differences between digitalPIRLS data and bridge data, in M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 12.1-12.33). Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb9281>

Retrouvez les travaux de la DEPP sur
education.gouv.fr/etudes-et-statistiques

Publications et archives

Retrouvez toutes les publications et archives de la DEPP sur
archives-statistiques-depp.education.gouv.fr

Jeux de données en open data

Retrouvez tous les jeux de données de la DEPP en open data sur
data.education.gouv.fr

Ministère de l'Éducation nationale

Direction de l'évaluation, de la prospective et de la performance (DEPP)