

LE RÉSULTAT AU TEST PISA EST-IL PRÉDICTIF DU SUCCÈS INDIVIDUEL ?

CHRISTIAN BELZIL
JORGEN HANSEN



Les rapports de projet sont destinés plus spécifiquement aux partenaires et à un public informé. Ils ne sont ni écrits à des fins de publication dans des revues scientifiques ni destinés à un public spécialisé, mais constituent un médium d'échange entre le monde de la recherche et le monde de la pratique.

Project Reports are specifically targeted to our partners and an informed readership. They are not intended for publication in academic journals nor aimed at a specialized readership, but are rather conceived as a medium of exchange between the research and practice worlds.

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du gouvernement du Québec, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO is a private non-profit organization incorporated under the Quebec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the government of Quebec, and grants and research mandates obtained by its research teams.

Les partenaires du CIRANO – CIRANO Partners

Partenaires corporatifs - Corporate Partners

Autorité des marchés financiers
Banque de développement du Canada
Banque du Canada
Banque Nationale du Canada
Bell Canada
BMO Groupe financier
Caisse de dépôt et placement du Québec
Énergir
Hydro-Québec
Intact Corporation Financière
Investissements PSP
Manuvie
Mouvement Desjardins
Power Corporation du Canada
Pratt & Whitney Canada
VIA Rail Canada

Partenaires gouvernementaux - Governmental partners

Ministère des Finances du Québec
Ministère de l'Économie, de l'Innovation et de l'Énergie
Innovation, Sciences et Développement Économique Canada
Ville de Montréal

Partenaires universitaires - University Partners

École de technologie supérieure
École nationale d'administration publique
HEC Montréal
Institut national de la recherche scientifique
Polytechnique Montréal
Université Concordia
Université de Montréal
Université de Sherbrooke
Université du Québec
Université du Québec à Montréal
Université Laval
Université McGill

Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.
CIRANO collaborates with many centres and university research chairs; list available on its website.

© Août 2026. Christian Belzil, Jorgen Hansen. Tous droits réservés. *All rights reserved.* Reproduction partielle permise avec citation du document source, incluant la notice ©. *Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.*

Les idées et les opinions émises dans cette publication sont sous l'unique responsabilité des auteurs et ne représentent pas les positions du CIRANO ou de ses partenaires. Le Ministre des Finances du Québec décline toute responsabilité à l'égard de l'information présentée dans le document, et que cette information ne reflète pas nécessairement les points de vue du Ministre. The observations and viewpoints expressed in this publication are the sole responsibility of the authors; they do not represent the positions of CIRANO or its partner The Quebec Minister of Finance assumes no responsibility for the information presented in this document, and such information does not necessarily reflect the views of the Minister.

Le résultat au test PISA est-il prédictif du succès individuel ?

Christian Belzil* et Jorgen Hansen†

Résumé

À partir de données longitudinales de l'Enquête auprès des jeunes en transition (EJET) conçue par Statistique Canada de concert avec le Programme international pour le suivi des élèves (PISA) de l'OCDE, nous analysons la capacité du test PISA à prédire les succès scolaires et professionnels ultérieurs.

De façon générale, nos résultats indiquent que les résultats obtenus dans les volets mathématique et lecture du test PISA sont beaucoup plus prédictifs du succès obtenu sur le marché du travail qu'ils ne le sont du succès académique ou de la probabilité d'obtenir un diplôme scientifique.

Nos résultats indiquent que l'OCDE est partiellement justifiée de considérer que le projet PISA peut mesurer des capacités individuelles à résoudre des problèmes concrets et pratiques, et non simplement des connaissances découlant directement des programmes scolaires. Cependant, étant donné la forte corrélation entre les résultats obtenus dans les deux volets, ainsi que la quasi-absence de différence entre leurs capacités prédictives, on ne peut exclure que les compétences mesurées par le programme PISA soient contaminées par un facteur latent commun assimilable à l'intelligence générale (une des critiques les plus communes adressées à la conception statistique du programme PISA).

Abstract

Using longitudinal data from the Youth in Transition Survey (YITS), designed by Statistics Canada in conjunction with the OECD's Programme for International Student Assessment (PISA), we examine the ability of PISA test scores to predict subsequent educational and labor market success.

Overall, our results indicate that performance on both components of the PISA assessment is much more predictive of success in the labor market than of academic success or the probability of obtaining a degree in a scientific field.

Our findings suggest that the OECD is partially justified in claiming that the PISA program measures individuals' ability to solve practical, real-world problems rather than simply assessing knowledge acquired directly through school curricula. However, given the strong correlation

* CREST-Institut Polytechnique de Paris, IZA et CIRANO

† Université Concordia, IZA et CIRANO

between performance on the two components of the assessment and the virtual absence of any difference in their predictive power, we cannot rule out the possibility that the competencies measured by the PISA program are influenced by a common latent factor akin to general intelligence—one of the most common criticisms directed at the statistical design of the PISA assessment.

Mots-clés/Keywords : formation, compétences / training, skills

Codes JEL/JEL Codes : I21, I24

Pour citer ce document / To quote this document

Belzil, C. et Hansen, J. (2026). Le résultat au test PISA est-il prédictif du succès individuel ? (2026RP-19, Rapport de projet, CIRANO).

<https://doi.org/10.54932/XIEB7789>

Sommaire

Le Programme international pour le suivi des acquis des élèves (PISA), qui évalue les performances des élèves en mathématiques, en sciences et en lecture dans divers pays depuis l'an 2000, est de loin la plus célèbre des évaluations internationales à grande échelle, connues sous le vocable de « *Large-Scale Assessments* » (LSA). En plus d'attirer l'attention des médias, ses résultats peuvent occasionnellement influencer les politiques publiques.

Dans le cadre des travaux de recherche effectués pour l'initiative « En avant math! », nous analysons la capacité du test PISA (tout particulièrement dans le domaine mathématique) à prévoir le succès individuel. Le but de ce rapport est de répondre aux deux questions suivantes :

- Les résultats individuels au test PISA, qui sont mesurés à l'âge de 15 ans, sont-ils capables de prédire les succès scolaires et professionnels ultérieurs?
- Dans quelle mesure les scores individuels PISA donnent-ils de l'information différente de celle déjà contenue dans les résultats scolaires contemporains de l'administration du test PISA et dans les mesures de traits non cognitifs?

Pour ce faire, nous utilisons les données longitudinales de l'Enquête auprès des jeunes en transition (EJET) conçue par Statistique Canada de concert avec le Programme international pour le suivi des élèves (PISA) de l'OCDE. L'EJET suit une cohorte d'étudiants nés en 1984. Ces données sont particulièrement instructives, puisque seuls deux pays (le Canada et le Danemark) permettent l'imbrication des scores individuels PISA dans des données observationnelles longitudinales.

Pour bien cerner la teneur des controverses entourant le programme PISA, le rapport débute par une présentation des perspectives historiques sur le développement des tests de réussite scolaire et des méthodes de mesure de l'intelligence, et montre ensuite les raisons pour lesquelles la validité statistique du test PISA est remise en question sérieusement par un grand nombre de chercheurs œuvrant en psychologie quantitative ou en statistique. Par la suite, le rapport présente les résultats de nos analyses permettant de relier les scores PISA, les résultats scolaires et les antécédents familiaux à diverses mesures de succès, telles que le nombre d'années d'études complétées, l'obtention d'un diplôme dans une filière scientifique, technologique, informatique ou mathématique (STIM) ainsi que les revenus à 30 ans.

De façon générale, notre approche est semi-structurelle. Nous obtenons les effets marginaux des scores et des résultats scolaires à l'aide de régressions linéaires estimées par la méthode des moindres carrés (sauf dans un cas où nous utilisons un modèle logit multinomial) et les interprétons de façon standard, c'est-à-dire comme le changement dans une mesure de succès (années de scolarité, etc.) induit par la variation d'un centile dans un score donné, toutes choses étant égales par ailleurs (antécédents familiaux ou autres).

Les résultats présentés dans notre rapport peuvent être synthétisés en sept éléments importants :

Résultat 1 : Il existe une corrélation forte et positive entre les mesures de réussite en lecture et en mathématiques, que ce soit pour PISA ou pour les résultats scolaires. De même, et pour un type de compétence donnée, il existe une corrélation positive, mais un peu moins forte, entre le score PISA et les résultats scolaires. Cela est globalement cohérent avec de nombreux résultats publiés

en psychologie quantitative indiquant qu'une forte proportion des habiletés spécifiques mesurées par PISA s'expliquent en fait par l'intelligence générale.

Résultat 2 : L'effet du facteur non cognitif sur les mesures de compétences (PISA ou résultats scolaires) est généralement inférieur à celui des facteurs cognitifs, mais il est plus marqué sur les résultats scolaires que sur les scores PISA.

Résultat 3 : Les scores individuels PISA et les résultats scolaires à 15 ans ont des effets limités, mais comparables, sur le nombre d'années d'études complétées, en particulier lorsqu'on les compare à l'impact du niveau d'éducation des parents. Une augmentation de 50 centiles (la différence entre le score médian et le premier centile) serait requise pour qu'un individu provenant d'une famille sans diplôme d'études secondaires puisse obtenir une année d'études supplémentaire (toutes choses étant égales par ailleurs).

Résultat 4 : Dans la population des individus nés en 1984, le score PISA en mathématiques fournit très peu d'information sur la probabilité d'obtenir un diplôme scientifique et a une capacité prédictive à peu près identique à celle des résultats scolaires. Nos résultats indiquent que l'étudiant appartenant au premier centile des scores PISA n'aurait qu'une probabilité de diplomation STIM supérieure de 0,05 à celle de l'étudiant médian dans la population.

Résultat 5 : Dans la sous-population des diplômés universitaires, les scores PISA en mathématiques sont encore moins discriminants que dans la population totale. Les résultats scolaires en mathématiques sont d'ailleurs deux fois plus importants (prédictifs) que le score PISA.

Résultat 6 : Alors que les scores PISA (en mathématiques aussi bien qu'en lecture) prédisent plus faiblement le succès scolaire ultérieur que les résultats scolaires à 15 ans, la relation est inversée lorsqu'on considère les revenus de travail à 30 ans, puisque les scores PISA ont un effet 3 fois plus grand que les résultats scolaires.

Résultat 7 : Pour la cohorte née en 1984, un diplôme scientifique (STIM) induit une prime de revenu de 54 % par rapport à l'ensemble des travailleurs sans diplôme universitaire, tandis qu'un diplôme obtenu dans une filière non scientifique induit une prime de 38 %. Cependant, le différentiel réel prévalant après la prise en compte des compétences mesurées à 15 ans est de 42 % pour un diplôme STIM et de 33 % pour les autres filières universitaires.

En conclusion, nos résultats portent un message sans doute moins négatif que celui qu'on retrouve dans une certaine littérature en psychologie quantitative (psychométrie). Bien qu'il soit sans doute vrai que les résultats individuels sont explicables en grande partie par le niveau d'habileté générale, leurs bonnes performances prédictives du succès sur le marché du travail indiquent que l'OCDE est tout au moins partiellement justifiée de considérer que le projet PISA peut mesurer des capacités individuelles à résoudre des problèmes concrets et pratiques, et non simplement des connaissances découlant directement des programmes scolaires. Néanmoins, on ne peut exclure que ces mêmes capacités soient assimilables à une autre facette, moins théorique et plus concrète, de l'intelligence individuelle.

Notre rapport se termine par deux recommandations concernant le design du test PISA qui permettraient de dissiper la confusion possible entre l'intelligence générale et les compétences spécifiques.

Premièrement, et afin de dissiper la confusion possible entre les effets de l'intelligence générale et les compétences spécifiques, il serait judicieux d'insérer un certain nombre d'items (questions) visant spécifiquement à mesurer l'intelligence générale afin qu'il soit possible de purger les résultats individuels de son effet. Cela pourrait se faire à l'aide de tests existants et internationalement reconnus (tests de Wechsler, des matrices de Raven, de Stanford-Binet, etc.)

Deuxièmement, et avec l'aide de ces items, il serait possible d'identifier les items du test PISA qui sont le moins affectés par l'intelligence générale, c'est-à-dire ceux qui reposeraient moins (qui « chargeraient » moins en langage statistique) sur les mesures de QI. Cela permettrait d'améliorer le design actuel en augmentant le nombre d'items indicatifs des facteurs ciblés spécifiquement par l'OCDE.

Table des matières

Sommaire	5
1- Introduction.....	10
2- Préambule historique	12
3- Que mesure le test PISA?	15
4- L'approche microéconométrique	19
5- Données et méthodologie.....	22
6- Les relations entre scores PISA, performances scolaires, traits non cognitifs et antécédents familiaux	25
Le rôle des traits non cognitifs.....	28
7- Le test PISA est-il prédictif du nombre d'années d'études?.....	30
Modélisation avec antécédents familiaux	30
Antécédents familiaux et robustesse	34
8- Le test PISA est-il prédictif du choix de filière?.....	37
La population totale	37
Sous-population des diplômés	39
9- Le test PISA est-il prédictif du niveau de revenu?	42
10- Les scores PISA et le biais d'habileté.....	45
11- Conclusions et implications pour le design du test PISA	47
Références.....	49

Liste des tableaux

Tableau 1 – Statistiques descriptives	24
Tableau 2 – Corrélations de rang : scores PISA, résultats scolaires et facteur non cognitif	25
Tableau 3 – Décomposition des scores PISA, des résultats scolaires et du facteur non cognitif .	26
Tableau 4 – La décomposition des scores PISA et des résultats scolaires par la méthode des moindres carrés : le rôle du facteur non cognitif	28
Tableau 5 – Décomposition du niveau d'éducation atteint par la méthode des moindres carrés avec antécédents familiaux	32
Tableau 6 – Décomposition du niveau d'éducation atteint par la méthode des moindres carrés sans antécédents familiaux.....	35
Tableau 7 – Modélisation du choix de filière avec un modèle logit multinomial : échantillon de population totale.....	38
Tableau 8 – Modélisation de la probabilité d'obtention d'un diplôme scientifique avec modèle de probabilité linéaire : échantillon des diplômés	40
Tableau 9 – Décomposition du log salaire (en 2015) : forme réduite	42
Tableau 10 – Biais d'habileté et rendements de l'éducation : années d'études et diplomation scientifique (STIM).....	46

1- Introduction

Dans la plupart des sociétés modernes, il est commun de s'appuyer sur des tests écrits pour sélectionner les individus, évaluer les élèves et les écoles et, même, mesurer des performances à l'échelle nationale. Le Programme international pour le suivi des acquis des élèves (PISA), qui évalue les performances des élèves âgés de 15 ans en mathématiques, en sciences et en lecture dans divers pays, en est l'exemple le plus célèbre. En plus d'attirer l'attention des médias, ses résultats peuvent même influencer les politiques publiques. Pour prendre un cas célèbre, il est généralement admis que les mauvais scores du test PISA rapportés en l'an 2000 ont incité l'Allemagne à réévaluer son système éducatif et à introduire une variété de réformes (Grek, 2009). Comme on le verra ci-dessous, l'interprétation des résultats du test PISA, et même sa pertinence, est au centre de nombreuses controverses de nature statistique ou méthodologique qui peuvent elles-mêmes remettre en question la validité des réactions des décideurs publics.

Considérant l'attention grandissante portée aux résultats nationaux dans les médias et dans le domaine public, il semble particulièrement pertinent de déterminer si l'importance attribuée à ces mêmes différences est validée par une analyse faite sur le plan individuel. En effet, dans la mesure où les concepteurs du test PISA prétendent pouvoir relier les différences de résultats entre pays aux différences de capacités des systèmes éducatifs à produire des compétences individuelles, il est normal de s'attendre à ce que les différences mesurées entre individus d'un même pays soient aussi indicatives du niveau de succès économique individuel ultérieur.

Dans le cadre des travaux de recherche effectués pour l'initiative « En avant math! », nous souhaitons nous pencher sur la capacité du test PISA (tout particulièrement dans le domaine des mathématiques) à prévoir le succès économique et professionnel individuel durant la phase initiale du cycle de vie. Un premier ensemble de questions se concentre sur l'interdépendance entre scores PISA, résultats scolaires et traits non cognitifs ainsi que leurs déterminants. Un second ensemble de questions vise à déterminer si les résultats individuels PISA sont susceptibles de prédire le succès individuel et à comparer leurs effets à ceux des résultats scolaires mesurés à 15 ans.

Dès le départ, il faut clarifier que, dans des conditions idéales, l'EJET donnerait accès au fichier original contenant tous les items utilisés et nous estimerions l'effet des facteurs latents identifiables par PISA sur le succès individuel de façon conjointe. Cela est malheureusement impossible, puisque nous n'avons accès qu'aux scores individuels, eux-mêmes générés à partir de plusieurs étapes décrites brièvement dans la section 3, ainsi qu'à l'ensemble des informations recueillies dans les différentes vagues d'entrevues. Nous utilisons donc ces mesures comme cela se fait souvent en économétrie appliquée, c'est-à-dire comme tout test standardisé (AFQT, test de QI, etc.), en tenant compte du fait que ces mesures individuelles n'ont pas la même précision que les moyennes nationales. Cependant, l'EJET ayant recueilli de l'information sur les résultats scolaires contemporains des scores PISA auprès des écoles, il nous est possible d'utiliser ces derniers comme une sorte de référence (*benchmark*) dans nos comparaisons.

Notre objectif est de répondre le plus précisément possible aux questions suivantes :

Les performances individuelles au test PISA :

i) sont-elles hautement corrélées aux résultats scolaires obtenus mesurés à âge égal?

- ii) sont-elles aussi prédictibles à partir des antécédents familiaux que le sont les résultats scolaires?
- iii) sont-elles plus affectées par les traits non cognitifs que les résultats scolaires?
- iv) ont-elles un impact plus fort que les résultats scolaires sur le nombre d'années d'études complétées?
- v) ont-elles un impact plus fort que les résultats scolaires sur la probabilité de compléter un diplôme scientifique (STIM)?
- vi) ont-elles un impact plus fort que les résultats scolaires sur les revenus à 30 ans?
- vii) Dans quelle mesure la prime salariale par année d'études est-elle expliquée par les performances individuelles au test PISA et les résultats scolaires?
- viii) Dans quelle mesure la prime salariale associée à un diplôme scientifique (STIM) est-elle expliquée par les performances individuelles au test PISA et les résultats scolaires?
- ix) Finalement, pour l'ensemble des mesures de succès analysées, les différences entre l'effet des scores PISA et des résultats scolaires sont-elles statistiquement significatives?

De façon générale, notre approche est semi-structurelle. Nous obtenons les effets marginaux des scores et des résultats scolaires à l'aide de régressions linéaires estimées par la méthode des moindres carrés (sauf dans un cas où nous utilisons un modèle logit multinomial) et les interprétons de façon standard, c'est-à-dire comme le changement dans une mesure de succès (années de scolarité, etc.) induit par la variation d'un centile dans un score donné, toutes choses étant égales par ailleurs (antécédents familiaux ou autres).

Pour bien cerner la teneur des controverses entourant le programme PISA, le rapport débute par une présentation des perspectives historiques sur le développement des tests de réussite scolaire et des méthodes de mesure de l'intelligence, et montre ensuite les raisons pour lesquelles la validité statistique du test PISA est remise en question sérieusement par un grand nombre de chercheurs en psychologie quantitative ou en statistique. Ces éléments se retrouvent dans les sections 2, 3 et 4.

Par la suite, le rapport présente les résultats de nos analyses permettant de relier les scores PISA, les résultats scolaires et les antécédents familiaux à diverses mesures de succès, telles que le nombre d'années d'études complétées, l'obtention d'un diplôme dans une filière scientifique, technologique, informatique ou mathématique (STIM) ainsi que les revenus à 30 ans.

Dans la conclusion du rapport, nous formulons deux recommandations qui, selon nous (et selon un grand nombre de chercheurs utilisant des méthodes quantitatives en psychologie), permettraient de dissiper la confusion possible entre les effets de l'intelligence générale et les compétences spécifiques.

2- Préambule historique

Pour bien comprendre la nature fondamentale du test PISA et la teneur des controverses qui l'entourent, il est nécessaire d'avoir une perspective historique sur le développement des tests de réussite scolaire ainsi que sur les méthodes de mesure de l'intelligence. Dans le texte qui suit, les termes « intelligence », « habileté générale » et « capacité générale » sont parfaitement interchangeables.

De façon théorique, l'intelligence est souvent définie comme la capacité à résoudre des problèmes, à comprendre des idées complexes et à penser de manière abstraite tout en apprenant de l'expérience et en s'adaptant au contexte dans lequel le raisonnement a lieu. Cette capacité individuelle est elle-même produite par une interaction entre le potentiel génétique de l'individu et son environnement (Hunt, 2010, p. 20). L'origine du concept de capacité cognitive générale est attribuée à Spearman (1904), qui a identifié un facteur d'intelligence générale à partir de données sur les performances des élèves dans différentes matières scolaires au moyen d'analyses factorielles. Le facteur d'intelligence générale, connu sous le vocable de « facteur g », est une variable latente qui explique les corrélations observées entre un ensemble diversifié de matières scolaires.

Afin de bien situer la nature réelle du test PISA dans l'univers des tests analysés en psychologie mathématique, en psychométrie, en économie et dans les sciences de l'éducation, il est indispensable de distinguer les tests de réussite scolaire des tests d'intelligence.

Les tests de réussite scolaire ont été élaborés à partir du milieu du XX^e siècle pour mesurer le concept de « connaissances générales ». Le but de ces tests était d'évaluer des compétences utiles à la réussite scolaire et à l'accomplissement de tâches moins scolaires. Ils visent à mesurer la performance des élèves dans différents domaines et permettent donc d'effectuer des comparaisons entre individus (pour un domaine précis) et entre domaines (pour un individu donné).

Les tests d'intelligence ont été conçus différemment puisque leur objectif principal était de mesurer l'aptitude générale sans égard à un domaine particulier. Dans un monde où les compétences spécifiques et l'intelligence générale seraient aisément séparables conceptuellement et statistiquement, les deux types de tests coexisteraient indépendamment et il n'existerait pas de confusion dans l'interprétation des facteurs mesurés par chacun de ces tests. En revanche, lorsque toute performance individuelle dans un domaine spécifique (mathématique, verbal ou autre) est affectée directement par l'intelligence générale, il devient beaucoup plus complexe de mesurer des compétences spécifiques, puisque les mesures de performance sont automatiquement contaminées par l'aptitude générale de chaque individu.

Si l'on part du principe que les aptitudes générales et spécifiques interagissent simultanément dans toutes les performances mesurant une compétence spécifique donnée, le degré de chevauchement entre les aptitudes spécifiques et générales devient un problème empirique de premier ordre. Sans quantifier leur importance relative, il devient très difficile d'évaluer des compétences spécifiques. Le degré de chevauchement pose non seulement un défi statistique complexe à résoudre, mais il génère aussi un certain nombre de controverses qui, comme on le verra ci-dessous, n'épargnent pas les débats entourant la pertinence du test PISA.

La cause principale de ces controverses est sans doute le fait qu'un des résultats les plus souvent reproduits en psychologie cognitive est l'identification d'un facteur d'aptitude générale, qui explique entre 50 % et 70 % de la variance des résultats scolaires des enfants, comme l'avait fait Spearman au début du 20^e siècle (Lubinski, 2004). Bien que le modèle du facteur « g » ait été critiqué et que d'autres modèles aient été proposés, des centaines d'études utilisant une variété de tests mentaux dans différentes populations ont également isolé des facteurs g (Gustafsson & Balke, 1993; Warne & Burningham, 2019). De plus, l'analyse des facteurs d'aptitude générale isolés à partir de différents tests de QI révèle que ces derniers sont fortement corrélés et qu'ils semblent donc tous mesurer un même facteur.

Malgré l'importance statistique du facteur d'aptitude général, de nombreux chercheurs en psychologie quantitative ont cherché à développer des modèles plus sophistiqués permettant la distinction entre intelligence générale et compétence spécifique. Un modèle important de l'intelligence humaine est celui à trois strates de Carroll (1993), comprenant un premier niveau inférieur composé de 50 à 60 habiletés étroitement définies et indépendantes, un deuxième niveau comprenant 8 à 10 habiletés larges et indépendantes et un niveau supérieur constitué d'un facteur unique d'aptitude intellectuelle générale, représenté par le facteur g de Spearman.

Le modèle factoriel de Cattell-Horn-Carroll (CHC) intègre le modèle de Carroll en ajoutant la distinction de Cattell entre intelligence fluide et intelligence cristallisée au niveau intermédiaire (Rindermann, 2007). L'intelligence « fluide » est définie comme la capacité intellectuelle indépendante des connaissances acquises, tandis que l'intelligence « cristallisée » mesure la capacité à utiliser les connaissances acquises. Cependant, d'un point de vue pratique, ce genre de modèle (en particulier celui à trois strates) apparaît difficile à estimer, puisqu'il requiert non seulement un nombre impressionnant d'items couvrant toutes ces différentes habiletés, mais aussi un nombre de facteurs latents tout aussi impressionnant.

Pour cette raison, les chercheurs ayant pour objectif d'estimer des modèles permettant de distinguer intelligence générale et habiletés spécifiques ont tendance à favoriser des modèles à deux niveaux. Cela explique sans doute l'intérêt grandissant suscité par le modèle bifactoriel (*bi-factor model* en anglais), qui est maintenant utilisé par de nombreux statisticiens utilisant les méthodes d'analyse factorielle basées sur des équations structurelles (*structural equation modeling*).

Si la popularité de l'approche bifactorielle est récente, le concept lui-même est ancien. En fait, Holzinger et Swineford (1937) ont développé la méthode bifactorielle en 1937. Celle-ci permet de décomposer un item de test observé en trois composantes, soit le facteur général, les facteurs spécifiques et l'erreur de mesure, afin d'évaluer l'importance relative des composantes générales et spécifiques. Les coefficients du modèle reflètent donc la mesure dans laquelle une performance intellectuelle est due à un trait général ou à une capacité plus spécifique.

Pour donner un exemple concret, si on appliquait le modèle bifactoriel à un ensemble d'items visant à évaluer les capacités en lecture et en mathématiques, toutes les mesures dépendraient d'un même facteur général et chaque item de type lecture dépendrait d'un facteur latent mesurant l'habileté spécifique en lecture, tandis que chaque item de type mathématique dépendrait d'un facteur latent représentant la compétence en mathématiques.

Pour un grand nombre de chercheurs en psychologie mathématique (statistique), seul un modèle bifactoriel permet d'évaluer des compétences spécifiques mesurées par divers tests, puisque ce dernier permet d'isoler (et donc d'éliminer) l'effet de l'intelligence individuelle des compétences spécifiques. Cette remarque s'applique aussi bien aux tests de réussite scolaire qu'au test PISA. Il est à noter qu'à notre connaissance, le modèle bifactoriel n'a (à ce jour) jamais été utilisé dans un modèle microéconométrique.

3- Que mesure le test PISA?

Le test PISA fait partie de l'ensemble des évaluations internationales à grande échelle du rendement des élèves, connues en anglais sous le vocable de « *Large-Scale Assessments* » (LSA). Il est de loin celui qui génère le plus d'intérêt médiatique et politique. Son importance s'explique sans doute à la fois par la promotion active que peut en faire l'OCDE ainsi que par la façon dont certains politiciens utilisent les résultats nationaux, en allant même jusqu'à induire des modifications dans divers programmes éducatifs.

Contrairement à beaucoup d'autres tests, PISA n'est pas lié directement à des connaissances et à des compétences spécifiques enseignées à l'école. Il vise plutôt à évaluer des compétences générales utiles dans la vie (Egelund, 2008). Une revue des documents officiels nous indique que les concepteurs ont tenu à différencier PISA de l'ensemble des tests de réussite scolaire en le caractérisant par la volonté de mesurer la manière dont les connaissances peuvent être utilisées dans divers contextes. D'ailleurs, les fonctionnaires du programme PISA utilisent le terme de « littératie » (*literacy* en anglais). PISA évalue donc la littératie dans trois domaines principaux : les mathématiques, la lecture et les sciences. Idéalement, ces diverses notions devraient mesurer « la capacité des élèves à appliquer des connaissances et des compétences dans des domaines et donc révéler des capacités à analyser et à raisonner dans des situations relativement concrètes » (OECD, 2007, p. 16)¹.

L'approche proposée par les concepteurs de PISA repose donc entièrement sur une vision multivariée dans laquelle l'intelligence générale ne joue aucun rôle. Malgré tout, il semble, à ce jour, n'y avoir aucune étude formelle, provenant de l'OCDE ou de chercheurs indépendants, démontrant que le test PISA mesure des compétences différentes de celles habituellement quantifiées dans des tests de réussite scolaire bien connus. Les fichiers individuels contiennent les données brutes et détaillées recueillies auprès des élèves, des écoles et des enseignants et incluent des informations contextuelles sur les élèves, leur foyer, leur environnement d'apprentissage et leurs attitudes, en plus des données sur leur performance. Ces fichiers servent de matériau de base pour l'analyse et sont utilisés pour produire les scores publiés. Les scores individuels obtenus à partir d'un modèle à facteur multivarié sont standardisés et utilisés pour les comparaisons internationales. Ces résultats sont ensuite interprétés comme indicateurs de la performance du système éducatif à l'échelle nationale.

Le processus de standardisation est complexe et sa description détaillée dépasse largement le cadre de notre étude. De façon schématique, le modèle de base appartient à la famille des modèles de réponse à l'item (*Item Response Theory*), dans lesquels on postule que la probabilité qu'un sujet donne une réponse correcte à un item est fonction de la compétence individuelle (traitée comme une variable latente) et de la difficulté de l'item (Rasch, 1960). Le niveau de compétence du sujet étant une valeur non observée (latente), il s'en suit que son estimation renvoie non pas à un seul paramètre, mais à une distribution. Conditionnellement aux réponses et aux caractéristiques de ce sujet, on peut inférer la distribution du paramètre de compétence d'un sujet ayant ces caractéristiques et ce patron de réponses aux items, puisque la distribution des compétences

¹ Occasionnellement, des domaines supplémentaires sont inclus, par exemple, la résolution de problèmes, la littératie financière et la pensée créative. Il est important de comprendre les résultats pour chaque dimension, qui sont rapportés séparément, conformément aux hypothèses fondamentales de l'étude selon lesquelles PISA mesure les domaines disciplinaires de manière fiable.

individuelles non observées et les réponses et caractéristiques individuelles observées sont reliées par le théorème de Bayes. PISA procède à plusieurs tirages dans cette distribution de valeurs dites « plausibles » et les utilisent afin de générer (simuler) différentes performances².

Il est de notoriété publique que la publication des rapports PISA entraîne fréquemment des débats universitaires et publics sur ce que ces tests mesurent vraiment et, donc, sur leur pertinence dans les discussions sur les politiques éducatives. Malgré l'existence de controverses dues au processus d'échantillonnage ou à divers aspects méthodologiques (calibration des paramètres de difficulté, incorporation des variables d'antécédents familiaux), la principale cause de polémique réside dans le fait que la notion de « littératie » utilisée par les fonctionnaires du programme PISA est très difficile à différencier de la notion d'intelligence générale, qui elle-même joue un rôle primordial en psychologie. PISA estime des facteurs latents de compétence (mathématiques, lecture, sciences) distincts, mais possiblement corrélés sous l'hypothèse implicite que le facteur d'aptitude générale (tel que le facteur *g*) est non pertinent pour les scores des élèves, puisque PISA considère que les performances individuelles ne sont imputables qu'aux habiletés spécifiques, qui sont elles-mêmes utilisées comme indicateurs de qualité des systèmes éducatifs. La question ultime est de savoir si la variance des éléments isolés contient bel et bien une proportion attribuable à un facteur général et, surtout, si celle-ci est importante (Brunner, 2008).

C'est justement sur cette base que de nombreuses études ont remis en question l'utilité des évaluations internationales à grande échelle, en particulier le programme PISA, en démontrant statistiquement que les tests PISA mesurent principalement la capacité cognitive générale plutôt que des compétences spécifiques à une discipline.

PISA mis à part, il est important de noter que de nombreuses études (menées dans différents pays) ont démontré que la capacité cognitive générale et les résultats scolaires sont très fortement corrélés. Pour citer quelques exemples, Walberg (1984) a calculé une corrélation moyenne de 0,71 entre diverses mesures de QI et la réussite scolaire. Kaufman et coll. (2012) ont plus récemment estimé des corrélations pouvant atteindre 0,77 à l'âge de 5 ans et à plus de 0,85 à 16 et 17 ans. Pour y parvenir, les auteurs ont estimé le facteur latent *g* (Spearman, 1904) ainsi qu'un autre facteur latent de capacité scolaire sous-jacent aux tests de lecture, de mathématiques et d'écriture.

Il existe d'ailleurs une méta-analyse intéressante (Zaboski, Kranzler & Gage, 2018) qui démontre de façon éloquente que les corrélations entre le facteur *g* et ceux caractérisant la lecture, la compréhension de texte et les mathématiques sont à peu près systématiquement supérieures à 0,7. Ce niveau de corrélation s'observe aussi dans des tests standardisés utilisés par diverses universités américaines, comme le *Scholastic Assessment Test* (SAT), l'*American College Readiness Assessment* (ACT) ou le célèbre *Armed Services Vocational Aptitude Battery* (ASVAB) de l'armée américaine, dont les mesures servent à la construction du célèbre *AFQT score*, souvent employé en microéconométrie. Par exemple, la corrélation entre le facteur *g* et le SAT est de l'ordre de 0,80 dans diverses études et tout aussi élevée pour les éléments du ASVAB (Frey, 2019; Koenig et coll., 2008).

² Okubo (2022) présente l'ensemble des concepts mathématiques et statistiques utilisés par PISA à différentes étapes du processus de normalisation.

En ce qui concerne PISA, selon Hopfenbeck et Kjærnsli (2018, p. 342), plus d'une cinquantaine d'articles ont étudié la relation entre l'intelligence et les scores PISA. Ils rapportent à peu près tous une corrélation globale très élevée (de l'ordre de 0,88) et concluent qu'en moyenne, environ plus de 80 % de la variation de PISA peut être attribuée aux capacités cognitives. À priori, on pourrait s'attendre à ce que l'importance de la capacité cognitive générale soit probablement plus marquée dans PISA que dans d'autres évaluations internationales à grande échelle, puisque ces dernières reposent en grande partie sur le contenu des programmes scolaires, tandis que PISA vise à évaluer dans quelle mesure les systèmes éducatifs forment des compétences générales utiles dans la vie plutôt que des connaissances et compétences spécifiques enseignées à l'école (Schleicher, 2007)³.

L'ensemble de ces résultats n'est guère étonnant puisque les corrélations entre les facteurs dérivés du modèle multidimensionnel PISA sont elles-mêmes très élevées, généralement entre 0,80 et 0,90 (Bond & Fox, 2001; Cromley, 2009). Pour certains pays portant sur l'évaluation PISA 2018 (OCDE, 2019), la corrélation moyenne entre mathématiques et lecture était de 0,80, de 0,82 entre mathématiques et sciences, et de 0,86 entre lecture et mathématiques. Lorsque ces facteurs sont fortement corrélés (lorsque les corrélations sont supérieures à 0,80), ils ont une validité discriminante faible et deviennent des mesures pratiquement non pertinentes des domaines distincts.

Ces résultats ont poussé certains chercheurs à appliquer le modèle bifactoriel aux données PISA afin d'évaluer dans quelle mesure les scores des élèves au test PISA reflètent la capacité cognitive générale plutôt que des compétences spécifiques à un domaine. Ce modèle est basé essentiellement sur le principe que le facteur général influe sur tous les items du test, parallèlement aux facteurs spécifiques que PISA étudie. Pokropek et coll. (2021, 2022) analysent les fichiers PISA allemand et polonais et concluent que les corrélations latentes entre les matrices progressives de Raven (une mesure de l'intelligence fluide) et les scores PISA étaient d'environ 0,73.

De façon synthétique, les auteurs montrent que les réponses des élèves aux items PISA reflètent la capacité générale plutôt que des compétences spécifiques et que le modèle bifactoriel génère des facteurs de capacité spécifiques plus purs, ayant une signification plus substantielle que les facteurs spécifiques issus du modèle multidimensionnel standard. Maximiser leur contenu spécifique au domaine, et donc leur fiabilité, nécessiterait des changements dans la sélection et la conception des items.

Certaines conclusions tirées des analyses des données PISA peuvent donc être infondées dans le cas où les scores des élèves issus des modèles multidimensionnels sont contaminés par la capacité générale. Par exemple, de nombreux décideurs publics préconisent des changements dans les programmes scolaires de mathématiques sur la seule base que le statut socio-économique et culturel de l'élève (*Economic, Social and Cultural Status* ou ESCS) est corrélé avec les scores PISA. Cependant, l'approche bifactorielle démontre aussi que le statut socio-économique et culturel de l'élève est davantage corrélé à la capacité générale qu'à la lecture ou à d'autres

³ Dans leur étude, Rindermann et Baumeister (2015) ont utilisé des évaluateurs experts afin d'examiner les capacités nécessaires pour répondre correctement aux items se trouvant dans PISA et à ceux de la TIMSS (*Trends in International Mathematics and Science Study*). Les auteurs ont constaté que l'intelligence générale et les connaissances générales jouaient un plus grand rôle pour PISA que pour TIMSS.

capacités spécifiques. Cela soulève donc des questions sur la conceptualisation du test PISA et l'interprétation de ses associations avec la performance des élèves (Schmidt & Burroughs, 2016). L'ESCS pouvant donc être expliqué par les capacités cognitives et les investissements précoces des parents, il s'en suit que les relations entre l'ESCS et les capacités spécifiques générées par le modèle multidimensionnel standard doivent être interprétées avec prudence.

On peut penser que si ces résultats étaient validés par d'autres études utilisant d'autres fichiers nationaux, ils auraient des implications majeures pour l'interprétation des résultats PISA. Ils indiqueraient que les analystes doivent tenir compte de la contamination des domaines spécifiques par la capacité cognitive générale lorsqu'ils considèrent les relations entre les scores de test et les variables socio-économiques, démographiques, scolaires et enseignantes. Les scores spécifiques aux domaines obtenus à partir des modèles bifactoriels, s'ils sont plus fiables, pourraient être plus significativement liés à des variables censées indiquer un niveau de compétence plus élevé.

Afin de maximiser la valeur informative du modèle bifactoriel, la littérature utilisant ce modèle prétend qu'il serait nécessaire de développer un groupe d'items différent et des critères distincts pour la sélection des items (Pokropek et coll., 2021). Lors de l'évaluation de la performance des élèves dans des domaines particuliers, le choix des items pourrait être guidé par leur charge sur les facteurs de capacité spécifiques plutôt que sur g. En revanche, pour évaluer les capacités générales de résolution de problèmes, le développement et le choix des items seraient guidés par leur charge sur g. Ces deux objectifs nécessitent de passer d'un cadre de modélisation multidimensionnel à un cadre bifactoriel.

4- L'approche microéconométrique

Une spécificité de la recherche en psychologie quantitative consacrée à la conception de tests de compétences est la validation statistique de ces tests à partir de variables mesurées dans un laps de temps relativement court après leur administration. La validation n'est pratiquement jamais basée sur des variables mesurant le succès individuel ultérieur (pendant le cycle de vie professionnel).

Les microéconomètres, qui modélisent les effets des habiletés cognitives et non cognitives sur le succès individuel, adoptent une approche différente. Pour bien comprendre l'apport des économistes, il faut partir du principe connu (mais souvent ignoré) selon lequel tout test de compétence ou de QI est ordinal par essence. L'absence de cardinalité requiert donc un besoin d'ancrage (*anchoring*) des tests de compétence ou des facteurs latents. L'ancrage de ces derniers dans un modèle économique permet en grande partie de relier leur échelle arbitraire à des variables mesurables de façon cardinale (probabilité d'obtenir un diplôme universitaire, fréquentation d'un programme scientifique, revenus individuels, etc.).

Dans la littérature microéconométrique récente, on commence à retrouver des évaluations de la valeur prédictive des divers types de tests. Dans leur article résumant les liens entre la psychologie de la personnalité et l'économie, Almlund et coll. (2011) concluent que les scores aux tests de réussite scolaire ne prédisent qu'une faible part de la variance du succès ultérieur dans la vie. Selon eux, les scores de ces tests à l'adolescence expliquent au plus de 15 % à 20 % de la variabilité des revenus à l'âge adulte.

Si l'on accepte ces résultats, il est fondamental de comprendre pourquoi beaucoup de tests de compétence perdent de leur pouvoir explicatif lorsqu'ils sont confrontés à des choix réels, puisque de prime abord, on pourrait être tenté de conclure simplement que ces tests mesurent des compétences non pertinentes sur le marché du travail. De façon schématique, on peut dire qu'il existe deux raisons fondamentales, qui sont elles-mêmes interreliées.

Premièrement, les performances mesurées dans le cadre des tests d'intelligence et de réussite scolaire dépendent, d'une certaine manière, des traits de personnalité et, tout particulièrement, du niveau de motivation de l'individu. Idéalement, tout test (que ce soit un test de QI ou un test de réussite) devrait être standardisé pour le niveau d'effort fourni. Cependant, cela est pratiquement impossible puisque l'effort est lui-même intrinsèquement non observable, en plus de n'avoir qu'une propriété ordinale.

La deuxième raison, elle-même reliée à la première, est que le niveau de performance peut aussi dépendre en grande partie du niveau d'incitation à la réussite, qu'il soit financier ou autre. Cela est spécialement vrai pour les individus obtenant des scores très bas dans divers tests. Selon Almlund et coll. (2011), de même que Borghans et coll. (2011), de nombreuses études montrent que l'élasticité des résultats aux incitations dépend souvent du niveau de « conscienciosité », qui constitue d'ailleurs un des traits du *Big Five*⁴. Certaines études montrent que, parmi les individus ayant des scores faibles à un premier test de QI, la performance aux tests ultérieurs peut être

⁴ Le concept de *Big Five* désigne l'ensemble des traits non cognitifs fondamentaux qui caractérisent chaque individu. La conscienciosité est le trait qui caractérise le plus la motivation individuelle à la réussite.

augmentée jusqu'à un écart-type en offrant des incitations, telles que de l'argent ou des bonbons, pour les bonnes réponses.

Ces problèmes sont aussi pertinents pour les tests de réussite scolaire, puisque ces derniers requièrent des connaissances acquises précédemment et que l'acquisition de connaissances est elle-même stimulée par la motivation, la curiosité et le niveau de persévérance.

La performance aux tests d'intelligence (et de réussite scolaire) dépend donc en partie des traits de personnalité du sujet, indépendamment de ses capacités cognitives, ainsi que de sa motivation à performer. Par exemple, un enfant intelligent incapable de rester calme pendant un examen ou peu motivé à fournir un effort peut obtenir de faibles scores à un test de QI. En d'autres termes, pour mesurer efficacement l'intelligence pure, il serait nécessaire de prendre en compte les incitations, la motivation ainsi que tout autre élément contextuel.

Les mêmes problèmes pertinents pour les tests de QI s'appliquent avec encore plus de force aux tests de réussite et aux notes scolaires. Les tests de réussite nécessitent des connaissances factuelles acquises par l'école et l'expérience de vie, qui sont en partie déterminées par la motivation, la curiosité et la persévérance du sujet. Cunha, Heckman et Schennach (2010) trouvent que les traits de personnalité facilitent l'acquisition de compétences cognitives mesurées par les tests de réussite. Ainsi, les traits de personnalité influencent indirectement les scores aux tests de réussite par le biais des connaissances accrues des individus possédant des niveaux élevés de certains traits de personnalité (Hansen, Heckman & Mullen, 2004; Heckman, Stixrud & Urzua, 2006).

Borghans et coll. (2011) examinent la pertinence de ces préoccupations. En utilisant deux sources de données des États-Unis et des Pays-Bas, ils établissent qu'une partie substantielle de la variance des notes et des tests de réussite est prédite par les mesures de personnalité. La contribution supplémentaire de la personnalité au-delà du QI est importante dans les deux ensembles de données. Les auteurs concluent que les notes et les scores aux tests de réussite prédisent mieux les résultats à l'âge adulte que le QI, car ils capturent également les traits de personnalité. Selon une autre interprétation, les connaissances acquises, mesurées par les tests de réussite et les notes, sont plus prédictives que l'intelligence fluide évaluée par le QI, précisément parce que les traits de personnalité affectent l'accumulation de connaissances.

De façon générale, et pour conclure cette section, la littérature microéconométrique récente illustre un problème fondamental d'identification en psychologie de la personnalité, soit que les comportements individuels sont influencés par les incitations et par des traits autres que ceux que les mesures cherchent à capturer. À ce stade, il est opportun de synthétiser trois enseignements majeurs provenant de la psychométrie et de l'économie qui nous permettent d'appréhender les capacités prédictives des performances individuelles à divers tests de compétences standardisés (y compris au test PISA), et celles des résultats scolaires.

Premièrement, de façon générale, la psychologie quantitative nous enseigne que les scores individuels PISA sont fort probablement beaucoup plus révélateurs des habiletés cognitives générales que des capacités spécifiques. Cela veut donc dire que l'effet mesuré des scores individuels sur des mesures de succès, telles que le niveau d'éducation atteint ou le niveau de

revenu, dépend indirectement de l'importance relative de l'intelligence générale et des habiletés spécifiques.

Deuxièmement, l'analyse économique nous indique que les performances individuelles mesurées pour toute tâche dépendent largement du niveau d'effort individuel lui-même induit par les incitations financières ou autres. Comme le niveau d'effort est lui-même impossible à mesurer, il s'en suit que l'effet mesuré des scores PISA et des résultats scolaires à 14 ans sur des mesures de succès individuels peut être contaminé par le niveau relatif d'effort fourni au test PISA par rapport aux résultats scolaires, mais aussi par l'importance relative des efforts individuels et des compétences (spécifiques ou générales) dans la détermination du niveau de succès individuel. Cela est précisément démontré dans Gneezy et coll. (2019), qui constatent que les élèves américains améliorent sensiblement leurs résultats lorsqu'ils sont incités, tandis que les élèves de Shanghai (déjà très performants) réagissent peu aux incitations, et concluent que les comparaisons internationales de scores (comme celles de PISA) peuvent être trompeuses.

Troisièmement, comme il est impossible de standardiser les résultats mesurés par rapport au niveau d'effort fourni, une solution imparfaite est de mesurer les traits non cognitifs des individus, puisque ces traits sont généralement corrélés avec la propension individuelle à fournir un degré élevé d'effort (pour un niveau d'incitation donné), et de les inclure dans tout modèle visant à évaluer l'impact des traits cognitifs sur un indicateur du succès individuel.

5- Données et méthodologie

Les résultats décrits ci-dessous ont été obtenus à partir des données de la cohorte A provenant de l'Enquête auprès des jeunes en transition (EJET), une étude longitudinale menée par Statistique Canada en collaboration avec Ressources humaines et Développement des compétences Canada, de concert avec le Programme international pour le suivi des élèves (PISA) de l'OCDE. L'enquête EJET ressemble, sur de nombreux aspects, au célèbre *National Longitudinal Survey of Youth* de 1997 (NLSY97).

Pour cette enquête, Statistique Canada a suivi une approche en deux étapes. À la première étape, un échantillon stratifié d'écoles a été sélectionné, afin d'assurer une couverture appropriée des 10 provinces canadiennes et fondée sur l'inscription des jeunes âgés de 15 ans à l'école l'année scolaire précédente. Comme préconisé par l'OCDE, les écoles pour personnes ayant des besoins spéciaux ont été exclues. Toutes ces exclusions représentent moins de 4 % des jeunes de 15 ans au Canada.

À la deuxième étape, un échantillon aléatoire simple d'étudiants a été sélectionné à l'intérieur des écoles. La population cible se composait d'étudiants inscrits dans un établissement d'enseignement au Canada le 31 décembre 1999 et âgés de 15 ans ce jour-là, c'est-à-dire qu'ils étaient nés en 1984. Les étudiants ont été rencontrés en entrevue initialement en avril ou en mai 2000, de nouveau de février à mai 2002 et de février à juin 2004, puis en 2008 et 2009.

L'enquête est conçue pour examiner les facteurs influençant la vie économique des jeunes, tels que l'éducation, la formation et le travail. Ces données incluent pratiquement toutes les expériences d'éducation formelle, la plupart des expériences sur le marché du travail, le rendement, les aspirations et les attentes, ainsi que les expériences d'emploi.

Une des spécificités des données EJET est l'imbrication d'informations individuelles issues du Programme international pour le suivi des acquis des élèves (PISA). Il nous est donc possible d'utiliser les informations provenant du test PISA et de les confronter avec d'autres mesures de compétences cognitives et non cognitives. À notre connaissance, seuls le Danemark et le Canada ont permis l'imbrication de résultats individuels du test PISA dans des données longitudinales permettant de relier les scores à des mesures de réussite éducative ou professionnelle observées pendant le cycle de vie.

L'évaluation PISA a été effectuée en 2000 et administrée dans la langue d'enseignement de l'école, soit l'anglais ou le français. Elle se composait de tests normalisés dans les domaines de la lecture, des mathématiques et des sciences. Tous les élèves ont été évalués en lecture, qui était le principal volet du test. La moitié des élèves ont également été évalués en mathématiques, et l'autre moitié, en sciences (sur la base d'un sous-échantillon aléatoire d'élèves du PISA dans les écoles). L'évaluation était fondée sur un test écrit de deux heures et, dans la composante « lecture », les élèves devaient exécuter une gamme de tâches portant sur différents types de textes, y compris repérer des renseignements particuliers, montrer qu'ils comprenaient le texte de façon générale, interpréter le texte et réfléchir sur le contenu et les caractéristiques du texte. Les textes comprenaient des passages en prose standard et divers types de documents, tels des listes, des formulaires, des graphiques et des diagrammes.

Comme dans le NLSY aux États-Unis, l'EJET comprend deux mesures de traits non cognitifs. La première est le sentiment de contrôle (ou de maîtrise) de sa propre vie (indice de Rotter). On a posé aux répondants sept questions évaluant dans quelle mesure ils croyaient pouvoir changer leur destinée ou leur succès dans la vie. Comme cela se fait couramment, une mesure scalaire a été initialement construite à partir des sept réponses.

La deuxième mesure est l'estime de soi (mesure de Rosenberg). On a posé aux répondants dix questions sur la façon dont ils se perçoivent. De nouveau, une mesure scalaire a été construite à partir de la moyenne des dix réponses. À partir de la moyenne de ces deux mesures, nous avons donc construit un indicateur unique (standardisé) que nous appelons le facteur non cognitif.

L'analyse présentée dans ce rapport utilise les variables suivantes : des indicateurs pour le niveau de scolarité le plus élevé obtenu par un des parents (secondaire, collégial, premier cycle universitaire, deuxième cycle universitaire ou plus) et un indicateur égal à 1 pour les adolescents vivant avec les deux parents biologiques au moment de l'enquête initiale. Afin de bien incorporer les différences géographiques potentielles, nous utilisons aussi l'information sur la région de résidence (provinces de l'Atlantique, Québec, Ontario, Manitoba-Saskatchewan, Alberta et Colombie-Britannique). À partir du questionnaire administré aux parents des étudiants en 2000, nous pouvons connaître le revenu total du ménage en 1999 (y compris les gains, le revenu de placements et le revenu sous forme de transferts gouvernementaux).

Un autre aspect intéressant de l'EJET est qu'elle contient des mesures de performance à l'école dans trois domaines spécifiques (lecture, mathématiques et sciences) ainsi qu'une mesure générale. Ces performances sont mesurées par un scalaire prenant des valeurs de 0 à 100. Évidemment, et contrairement aux scores PISA, les résultats scolaires ne sont pas standardisés puisqu'ils sont probablement influencés en partie par les différences de qualité entre les écoles.

Le tableau 1 contient les statistiques descriptives pour l'ensemble des variables utilisées. Ces statistiques sont calculées à partir de l'échantillon total provenant de la cohorte A de l'EJET. Au vu des objectifs principaux de notre étude, le critère de sélection principal vise à obtenir la sous-population des jeunes de 15 ans pour qui un score en mathématiques et en lecture est disponible. À ce critère s'ajoute simplement une condition d'observabilité de chaque variable, ce qui explique les différences de nombre d'observations entre chaque variable au tableau 1.

Dans le tableau, les scores PISA en mathématiques, lecture et sciences sont dénotés PISA_M, PISA_L et PISA_S. Les résultats scolaires en mathématiques, lecture et sciences sont dénotés GPA_M, GPA_L et GPA_S. Le résultat scolaire général (toutes matières confondues) est dénoté GPA_T. Finalement, le facteur non cognitif apparaissant à partir du tableau 2, et dénoté NC, est lui-même construit à partir des composantes « estime de soi » et « efficacité » apparaissant au tableau 1.

Les scores moyens PISA des jeunes de notre échantillon (534 pour le verbal et 530 pour les mathématiques) sont pratiquement identiques aux scores officiels publiés par l'OCDE pour le Canada en 2000 (534 pour le chapitre verbal et 533 pour la composante mathématiques). Ces scores sont aussi supérieurs à la moyenne des pays participants, puisque la moyenne (tous pays confondus) est fixée à 500 par l'OCDE.

On note aussi que 15 % des jeunes vivent dans des familles dont au moins un parent est détenteur d'un diplôme universitaire de premier cycle, tandis que 6 % vivent avec un parent ayant un diplôme de deuxième cycle. En ce qui concerne la composition familiale, on remarque qu'un peu plus de 70 % des jeunes déclarent vivre avec leurs parents biologiques. Finalement, le revenu moyen des ménages en 1999 se situait juste en dessous de 70 000 \$ (69 400 \$).

Tableau 1 – Statistiques descriptives

	Moyenne	Écart-type	N ^{bre} d'observations
Mesures cognitives			
PISA-M (mathématiques)	529,8	86,2	14 446
PISA-L (lecture)	534,4	98,1	14 463
GPA-M (mathématiques)	73,3	14,3	13 312
GPA-L (lecture)	74,9	12,6	13 402
GPA-S (sciences)	75,1	13,4	13 345
GPA-T (toutes les matières)	75,8	11,2	13 354
Mesures non cognitives			
Estime de soi	-0,00	1,00	14 040
Efficacité	0,00	1,00	14 040
Antécédents			
P. Univ_2ièmeC	0,06	-	14 038
P. Univ_1erC	0,15	-	14 038
P. Collégial	0,40	-	14 038
P. Secondaire	0,27	-	14 038
P. Sans diplôme	0,12	-	14 038
Famille nucléaire	0,71	0,44	14 460
Fille	0,50	-	14 482
Revenu des parents	69,4	52,1	14 485
Ouest	0,31	-	14 485
Ontario	0,37	-	14 485
Québec	0,23	-	14 485
Atlantique	0,08	-	14 485

Remarques :

1. Les scores PISA en mathématiques, lecture et sciences sont dénotés PISA_M, PISA_L et PISA_S. Les résultats scolaires en mathématiques, lecture et sciences sont dénotés GPA_M, GPA_L, et GPA_S. Le résultat scolaire total (toutes matières confondues) est dénoté GPA_T.

2. Les variables P. Univ_2ièmeC, P. Univ_1erC, P. Collégial, P-Secondaire et P-Sans diplôme sont des indicateurs binaires du niveau d'éducation des parents. La variable « Famille nucléaire » est égale à 1 pour les individus vivant avec leur famille nucléaire au moment de l'enquête. Le revenu des parents est exprimé en milliers de dollars de 2000.

3. Les statistiques sont mesurées pour l'an 2000 et sont basées sur l'échantillon pondéré.

6- Les relations entre scores PISA, performances scolaires, traits non cognitifs et antécédents familiaux

Comme cela a été mentionné dans la section 2, toutes ces mesures provenant de divers tests de compétence, de performance ou simplement d'auto-évaluation n'ont qu'une valeur ordinale, puisque leurs échelles sont purement arbitraires. Pour cette raison, et lorsque ces dernières apparaîtront comme variable explicative dans certaines de nos analyses, nous utiliserons le rang individuel exprimé en centile.

Corrélations entre mesures de compétences et antécédents familiaux

Dans un premier temps, nous avons mesuré les corrélations de rang de même que les corrélations obtenues à partir des mesures brutes. Les résultats étant comparables, nos commentaires se basent uniquement sur les corrélations de rang qui sont présentées dans le tableau 2.

Tableau 2 – Corrélations de rang : scores PISA, résultats scolaires et facteur non cognitif

	PISA_M	PISA_L	GPA_M	GPA_L	GPA_S	GPA_T	N-C
PISA-M	1,00	-	-	-	-	-	-
PISA-L	0,66	1,00	-	-	-	-	-
GPA-M	0,40	0,36	1,00	-	-	-	-
GPA-L	0,33	0,44	0,43	1,00	-	-	-
GPA_S	0,42	0,44	0,58	0,55	1,00	-	-
GPA-T	0,45	0,50	0,69	0,69	0,74	1,00	-
NC	0,29	0,32	0,33	0,31	0,22	0,41	1,00

Remarques :

1. Les statistiques sont mesurées pour l'an 2000 et sont basées sur l'échantillon pondéré. Tous les scores sont mesurés en rang centile.
2. Les scores PISA en mathématiques, lecture et sciences sont dénotés PISA_M, PISA_L et PISA_S. Les résultats scolaires en mathématiques, lecture et sciences sont dénotés GPA_M, GPA_L, et GPA_S. Le résultat scolaire total (toutes matières confondues) est dénoté GPA_T.
3. Le facteur non cognitif (NC) est une moyenne de la mesure d'estime de soi ainsi que de l'index de Rotter, qui mesure le degré de confiance de l'individu.

Premièrement, et comme rapporté dans de nombreuses études citées dans la section 3, et utilisant les données brutes du programme PISA, la corrélation entre la mesure de performance en lecture et en mathématiques égale à 0,66 est positive et très élevée. Ce niveau de corrélation est compatible avec l'hypothèse souvent vérifiée que les performances individuelles dans chacun des domaines spécifiques fortement corrélés sont générées principalement par un facteur commun interprétable, comme le niveau d'intelligence générale.

Deuxièmement, nous posons la question suivante : dans quelle mesure les scores PISA individuels sont-ils expliqués par les antécédents familiaux, tels que l'éducation des parents ou la composition familiale?

Pour répondre à cette question, nous examinons les résultats d'une régression expliquant le rang individuel (la variable dépendante) en fonction de l'ensemble des caractéristiques qui mesurent les

antécédents familiaux (éducation des parents, genre, composition familiale et région). Les résultats se trouvent dans le tableau 3.

Tableau 3 – Décomposition des scores PISA, des résultats scolaires et du facteur non cognitif

Var. dépendantes	PISA-M	PISA-L	GPA-M	GPA-L	GPA-S	NC
	Paramètres estimés					
Constante	42,64**	32,76**	31,92**	17,31**	28,47**	52,86**
P. Univ_2ièmeC	26,95**	30,20**	17,35**	19,00**	18,25**	14,51**
P. Univ_1erC	21,85**	23,30**	13,22**	15,53**	14,28**	11,34**
P. Collégial	13,28**	15,58**	6,10**	7,87**	6,52**	5,23**
P. Secondaire	9,43**	10,74**	3,36**	4,85**	4,31**	3,54**
Famille nucléaire	5,01**	3,32**	4,15**	5,65**	5,35**	2,65**
Fille	-2,43**	9,83**	2,59**	11,84**	3,64**	-6,00**
Revenu des parents	0,02**	0,02*	-0,01	0,01	0,01	0,01
Ouest	-5,38**	-1,52*	-3,93**	1,71**	-3,15**	-8,19**
Ontario	-11,57**	-5,02**	-3,32**	0,15	-4,49**	-6,21**
Atlantique	-13,51**	-9,21**	0,53	7,55**	0,23	-9,37**
Québec	-	-	-	-	-	-
R carré	0,10	0,12	0,04	0,12	0,06	0,04
N ^{bre} d'observations	13 976	13 992	12 880	12 968	12 921	13 448

Remarques :

1. Les résultats sont basés sur l'échantillon pondéré. Tous les scores sont mesurés en rang centile.
2. Les estimés avec « ** » (*) sont significatifs à 1 % (5 %).
3. Les variables P. Univ_2ièmeC, P. Univ_1erC, P. Collégial et P-Secondaire sont des indicateurs binaires du niveau d'éducation des parents. La variable « Famille nucléaire » est égale à 1 pour les individus vivant avec leur famille nucléaire au moment de l'enquête. Le revenu des parents est exprimé en milliers de dollars de 2000.
4. Le facteur non cognitif (NC) est une moyenne de la mesure d'estime de soi ainsi que de l'index de Rotter, qui mesure le degré de confiance de l'individu.

De façon générale, et comme révélé par les différents R carrés, les antécédents familiaux expliquent entre 10 % (pour les résultats en mathématiques) et 12 % (le volet verbal-lecture) de la variance des performances individuelles. Dans les deux cas, on note une forte corrélation avec le niveau d'éducation des parents. Par exemple, les enfants dont au moins un parent a obtenu un diplôme universitaire de 1^{er} cycle obtiennent un rang en mathématiques d'environ 22 centiles supérieur à celui des individus dont les parents n'ont pas obtenu de diplôme d'études secondaires. Pour le volet lecture, l'avantage est de l'ordre de 23 centiles. Pour ceux dont les parents ont obtenu un diplôme universitaire de deuxième cycle (ou plus), les différentiels sont encore plus élevés (27 centiles en mathématiques et 30 centiles en lecture).

Bien que le revenu des parents semble aussi augmenter les scores PISA, son effet est beaucoup plus faible et n'est pas très significatif. Dans les deux cas, l'effet associé à une variation de 1 000 dollars est approximativement égal à 0,02 et nous indique donc qu'une différence de 100 000 dollars, équivalente à deux fois l'écart-type des revenus familiaux dans la population,

induirait seulement une différence de 2 centiles. Sans surprise, les résultats montrent que les filles dominent les garçons de 10 centiles dans le volet verbal-lecture. Cependant, elles ont en même temps un déficit d'environ 2 centiles en mathématiques.

Finalement, il est intéressant de noter certaines différences géographiques. Le Québec, qui est la région de référence dans la régression, est la province qui performe le mieux dans les deux volets du test PISA. Cela est particulièrement marqué en mathématiques, où les étudiants québécois dominent les Ontariens de 12 centiles (toutes choses étant égales par ailleurs). Cette différence est très précisément estimée.

Comme l'un de nos objectifs est ultimement de comparer la capacité prédictive des scores PISA avec celle des résultats scolaires et celle des traits non cognitifs, il est aussi pertinent de faire le même exercice avec ceux-ci. Comme cela a été noté avec les résultats individuels PISA, les résultats scolaires individuels sont aussi fortement corrélés de façon positive. Par exemple, la corrélation entre les résultats en mathématiques et les résultats en lecture, égale à 0,44, est moins élevée que la valeur mesurée entre le score PISA en mathématiques et le score en lecture, mais celle entre les résultats scolaires en mathématiques et ceux en sciences est plus élevée (0,58). De même, la corrélation entre les résultats en sciences et ceux en lecture est de l'ordre de 0,60.

Non seulement les corrélations entre les scores obtenus en mathématiques et en lecture sont-elles élevées (que ce soit entre les scores PISA ou entre les résultats scolaires), mais il existe aussi des corrélations positives et fortes entre les scores PISA et les résultats scolaires. Comme mentionné plus tôt, une forte corrélation est compatible avec les thèses soutenues par de nombreux chercheurs actifs dans la littérature en psychologie quantitative. Par exemple, pour la mesure la plus générale des résultats scolaires (GPA-T), on obtient une corrélation de 0,50 avec le score PISA en lecture et de 0,45 avec le score PISA en mathématiques.

De façon générale, on peut donc dire que, tout comme le sont les scores individuels PISA, les résultats scolaires observés dans des domaines spécifiques s'expliquent fort probablement en grande partie par le niveau d'habileté générale individuelle, tel que le facteur g des psychologues.

La proximité des résultats scolaires avec les scores PISA ne s'arrête pas uniquement aux corrélations. Comme cela est illustré dans le tableau 3 (colonnes 3, 4 et 5), les antécédents familiaux expliquent entre 4 % et 12 % de la variance des résultats scolaires. Les résultats scolaires en lecture sont ceux qui sont le plus expliqués par les antécédents familiaux, tandis que ces derniers jouent un moindre rôle pour les résultats en mathématiques et en sciences.

D'ailleurs, comme pour PISA, le niveau d'éducation des parents est un facteur crucial, mais il semble moins discriminant que pour les scores PISA. Par exemple, la différence de rang expliquée par le différentiel de premier cycle universitaire et non-diplômé est de l'ordre de 13 centiles pour les mathématiques et de 15 centiles pour la compétence verbale. Cependant, la différence entre garçons et filles est de loin la plus spectaculaire. Les filles ont un rang moyen en lecture excédant celui des garçons de 12 centiles. Elles dominent aussi les garçons en mathématiques de 3 centiles et de 5 centiles en sciences.

Il est donc à noter que la différence de performance en mathématiques entre les garçons et les filles semble dépendre de la nature de la mesure, puisque les garçons dominent dans le test standardisé (PISA), mais réussissent moins bien que les filles en classe.

Le rôle des traits non cognitifs

Comme cela a été mentionné en introduction, certains économistes considèrent que la capacité prédictive des mesures obtenues lors de tests de compétences peut dépendre des incitations à la réussite et donc que le niveau de réaction aux incitations dépend lui-même des traits non cognitifs de l'individu. De façon générale, les traits non cognitifs englobent toutes les caractéristiques individuelles pouvant expliquer les performances dans des tâches cognitives, en sus des habiletés cognitives pures. Par exemple, le niveau de motivation et de conscienciosité d'un individu est susceptible d'affecter ses performances.

Dans le contexte présent, où les scores PISA reflètent l'effort ponctuel que l'étudiant a bien voulu consentir ainsi que son niveau de motivation, il est probable que les traits non cognitifs puissent affecter différemment les scores PISA obtenus et les résultats scolaires puisque ces derniers représentent un enjeu réel pour la réussite professionnelle ultérieure (Heckman, Jagelka & Kautz, 2019). Si tel est le cas, des régressions simples des scores PISA et des résultats scolaires sur le facteur cognitif peuvent nous éclairer sur ce phénomène. Afin de contrôler le niveau d'habileté générale affectant les scores PISA, nous incluons aussi le GPA total dans certaines spécifications expliquant les scores PISA_M et PISA_L.

Tableau 4 – La décomposition des scores PISA et des résultats scolaires par la méthode des moindres carrés : le rôle du facteur non cognitif

Var. dépend.	PISA-M (math)			PISA-L (lecture)			GPA total
Constante	29,87**	34,30**	36,73**	29,31**	33,52**	37,15**	17,90**
GPA total	0,44**	0,50**	-	0,49**	0,56**	-	-
Non cognitif	0,13**	-	0,30**	0,13**	-	0,31**	0,36**
R carré	0,21	0,20	0,09	0,27	0,25	0,10	0,17
Nbre d'observations	13 324			13 913			13 079

Remarques :

1. Les résultats sont basés sur des régressions obtenues à l'aide de l'échantillon pondéré.
2. Tous les scores sont mesurés en rang centile.
3. Les estimés avec « ** » (*) sont significatifs à 1 % (5 %).

Les résultats des régressions se trouvant au tableau 4 illustrent assez clairement comment les traits non cognitifs peuvent affecter différemment les scores PISA et les résultats scolaires. Les paramètres estimés indiquent précisément comment une augmentation du facteur non cognitif d'un centile affecte le rang (en centile) des deux scores PISA ainsi que les résultats scolaires généraux.

Les résultats indiquent globalement une importance plus grande des traits non cognitifs sur les résultats scolaires que sur les scores PISA. Les paramètres estimés, 0,30 et 0,31 (colonnes 3 et 6) pour les scores PISA et 0,36 (colonne 7) pour les résultats scolaires, indiquent qu'une variation

positive de 10 centiles du facteur non cognitif se traduit par une augmentation des scores PISA de 3 centiles, mais de près de 4 centiles pour les résultats scolaires. Si l'on approxime une variation d'un écart-type par une différence de 35 centiles (ce qui serait le cas pour une loi normale), cette différence correspondait à une augmentation de 10 centiles dans les deux scores PISA et de 14 centiles pour les résultats scolaires. L'effet du facteur non cognitif sur les résultats scolaires est donc d'environ 40 % supérieur à son effet sur les scores PISA.

Ces résultats sont donc cohérents à la fois avec le postulat que les mesures de compétences sont partiellement expliquées par des traits de personnalité autres que les seuls traits cognitifs et avec la thèse défendue par certains économistes que l'effet des traits de personnalité est plus marqué sur les mesures de succès individuels pertinents dans la vie professionnelle que sur les performances individuelles à divers tests standardisés (Heckman, Jagelka & Kautz, 2019).

Ces résultats impliquent donc que, pour le reste de nos analyses visant à comparer les propriétés prédictives des scores PISA, il serait approprié de prendre en compte le facteur cognitif et de l'incorporer dans nos estimations.

De cette section, nous retenons deux premiers résultats importants :

Résultat 1 : Il existe une corrélation forte et positive entre les mesures de réussite en lecture et en mathématiques, que ce soit pour PISA ou pour les résultats scolaires. De même, et pour un type de compétence donnée, il existe une corrélation positive, mais un peu moins forte, entre le score PISA et les résultats scolaires. Ce résultat est globalement cohérent avec de nombreux résultats publiés en psychologie quantitative indiquant qu'une forte proportion des habiletés spécifiques mesurées par PISA est expliquée par l'intelligence générale.

Résultat 2 : L'effet du facteur non cognitif sur les mesures de compétences (PISA ou résultats scolaires) est généralement inférieur à celui des facteurs cognitifs, mais son effet est plus marqué sur les résultats scolaires que sur les scores PISA.

7- Le test PISA est-il prédictif du nombre d'années d'études?

L'objectif principal de l'exercice synthétisé dans cette section est de déterminer si les résultats individuels au test PISA ont une capacité prédictive du niveau d'éducation obtenu, et si celle-ci est plus grande que la capacité prédictive d'une mesure classique de succès scolaire en classe (au même âge) ou même que celle d'une mesure de trait non cognitif. En d'autres termes, le test standardisé PISA donne-t-il plus ou moins d'information que les résultats scolaires, qui, eux, ne sont pas standardisés?

Pour répondre à cette question, nous utilisons l'accès aux mesures spécifiques à chaque domaine (mathématiques et lecture) puisque, pour chacun de ces domaines, nous avons à la fois le résultat PISA ainsi qu'une mesure de succès scolaire. Cela nous permet donc de confronter chacun des deux scores PISA à sa mesure de succès scolaire correspondante.

Il est important de noter que l'information requise pour notre analyse a été obtenue à partir des enquêtes de suivi décrites dans la section « Données et méthodologie » et que cela explique la diminution du nombre d'observations à partir du tableau 5.

Modélisation avec antécédents familiaux

Dans un premier temps, nous suivons l'approche classique souvent utilisée en économie de l'éducation, qui consiste à mesurer l'éducation par le nombre d'années d'études. Pour ce faire, nous régressons donc le niveau de scolarité atteint (mesuré en années) sur l'ensemble des variables explicatives indiquant les antécédents familiaux (éducation des parents, composition familiale, revenu, genre, régions), le score PISA en mathématiques (PISA_M), la mesure de succès scolaire en mathématiques (GPA_M), ainsi que la mesure moyenne des deux traits non cognitifs décrits précédemment (NC).

Il est important de noter que les effets des scores PISA et des résultats scolaires sur le niveau d'éducation que nous mesurons doivent être interprétés comme étant des effets nets purgés de l'effet des antécédents familiaux, tels que le niveau d'éducation des parents et la composition familiale, puisque notre spécification incorpore à la fois les effets des antécédents et des compétences mesurées à 15 ans. Toutes ces variables étant cependant corrélées avec les scores PISA et les résultats scolaires, une analyse des résultats obtenus lorsque les antécédents familiaux sont exclus de la régression sera présentée ci-dessous.

Comme nous utilisons le rang des résultats individuels comme variable explicative, leurs effets marginaux respectifs sont très faciles à interpréter. Ils sont parfaitement comparables entre diverses mesures de compétences ou entre différents types de mesure pour une compétence donnée, et peuvent donc être facilement utilisés afin de comparer leur valeur prédictive.

Notre stratégie étant de comparer les scores PISA et les résultats scolaires pour chaque compétence donnée, le même exercice est répliqué une première fois en remplaçant les mathématiques par la compétence en lecture. Pour chacune de ces compétences, nous présentons les résultats de trois régressions : une où le test PISA et le résultat scolaire sont inclus ainsi que deux autres où chacune

de ces mesures est incluse séparément. Dans tous les cas, les antécédents familiaux et le facteur non cognitif sont inclus dans la régression⁵.

Pour chacun de ces groupes de régressions, il est facile de procéder à un test statistique dont l'hypothèse nulle est l'égalité des effets marginaux des mesures PISA et des mesures scolaires. Une égalité de ces effets indiquerait aussi, jusqu'à un certain degré, que les deux scores mesurent un même facteur latent, bien qu'on ne puisse exclure qu'il existe d'autres variables d'intérêt pour lesquelles les effets sont différents.

Les résultats obtenus pour la compétence en mathématiques se retrouvent dans les trois premières colonnes du tableau 5, tandis que ceux obtenus pour la compétence en lecture sont dans les colonnes 4, 5 et 6. Les résultats de la première colonne, aussi bien qu'une comparaison entre ceux des colonnes 2 et 3 dans lesquels les deux mesures sont incluses séparément, indiquent clairement que l'indicateur de réussite scolaire en mathématiques (le GPA-M) donne à peu près la même information que le score PISA. Précisément, lorsque les deux mesures sont incluses, une différence de 1 centile dans le score PISA ou dans le GPA-M correspond à une même augmentation d'environ 0,018 année de scolarité. Ce résultat est confirmé par la valeur du test statistique F égal à 0,34 (avec une valeur P égale à 0,56).

Pour rappel, la valeur P (en anglais *p-value*, pour « *probability value* »), aussi connue sous le vocable de « probabilité critique », est la probabilité d'observer (sous l'hypothèse nulle) une valeur statistique au moins aussi extrême que celle observée. En pratique, plus la valeur P est petite, plus on peut rejeter l'hypothèse nulle avec confiance. Dans ce cas, avec une valeur P égale à 0,56, il est donc inapproprié de rejeter l'hypothèse nulle, et on peut donc conclure que ces deux mesures de compétences en mathématiques ont des effets identiques sur le niveau de scolarité. Cela indique donc l'existence d'une forte similitude entre ces deux mesures.

À priori, il est difficile de déterminer si un effet rang-rang de 0,018 indique un fort niveau de prédictibilité sans le comparer aux effets estimés pour les antécédents familiaux, et plus particulièrement le niveau d'éducation des parents. Cependant, le score PISA et l'éducation des parents ayant des échelles complètement différentes, il n'existe pas de méthode évidente.

Afin d'illustrer tangiblement le niveau de prédictibilité des scores PISA et des résultats scolaires sur le nombre d'années d'études, nous proposons la comparaison suivante. Premièrement, nous référons aux résultats du tableau 3 (1^{re} colonne) pour obtenir une variation de rang dans le score PISA associée à la différence maximale d'éducation des parents, c'est-à-dire la différence moyenne de rang entre les jeunes provenant de familles sans diplôme du secondaire et ceux dont les parents ont un diplôme universitaire. Cette différence est d'environ 23 centiles pour les mathématiques aussi bien que pour la compétence en lecture.

⁵ Cet exercice a aussi été répliqué une deuxième fois en utilisant la moyenne des deux scores PISA et en la confrontant aux résultats scolaires tous domaines confondus (le GPA_T). Cependant, comme les résultats obtenus dans ce cas général sont souvent identiques aux deux autres, nous ne les présentons pas afin d'alléger les tableaux de résultats.

Tableau 5 – Décomposition du niveau d'éducation atteint par la méthode des moindres carrés avec antécédents familiaux

Var. dépendante	Années d'études					
	(1)	(2)	(3)	(4)	(5)	(6)
PISA-M (Math)	0,014**	0,018**	-	-	-	-
PISA-L (Lecture)	-	-	-	0,016**	0,020**	-
GPA-M (Math)	0,011**	-	0,017**	-	-	-
GPA-L (Lecture)	-	-	-	0,013**	-	0,019**
NC (non cognitif)	0,008**	0,011**	0,009**	0,007**	0,009**	0,010**
P. Univ_2ièmeC	1,036**	1,032**	1,373**	0,881**	1,001**	1,218**
P. Univ_1erC	0,945**	1,026**	1,192**	0,849**	1,017**	1,114**
P. Collégial	0,759**	0,735**	0,901**	0,647**	0,669**	0,854**
P. Secondaire	0,524**	0,485**	0,654**	0,446**	0,465**	0,583**
Famille nucléaire	0,318**	0,321**	0,372**	0,315**	0,379**	0,315**
Filles	0,876**	0,951**	0,811**	0,562**	0,683**	0,661**
Revenu des parents	0,003**	0,003**	0,003**	0,003*	0,003*	0,003*
Ouest	0,186	0,283**	0,129	0,079	0,191	0,067
Ontario	0,528**	0,624**	0,348**	0,466**	0,564**	0,313*
Atlantique	0,705**	0,828**	0,507**	0,571**	0,735**	0,397**
Québec	-	-	-	-	-	-
Constante	10,879**	10,838**	11,249**	11,091**	11,014**	11,410**
R carré	0,28	0,27	0,25	0,29	0,27	0,25
N ^{bre} d'observations	1752	1857	1754	1763	1857	1765
Stat. F	0,34	-	-	0,59	-	-
Valeur P	0,56	-	-	0,44	-	-

Remarques :

1. Les résultats sont basés sur des régressions obtenues à l'aide de l'échantillon pondéré.
2. Le niveau d'éducation atteint est mesuré en nombre d'années d'études complétées.
3. Les variables P. Univ_2ièmeC, P. Univ_1erC, P. Collégial et P-Secondaire sont des indicateurs binaires du niveau d'éducation des parents. La variable « Famille nucléaire » est égale à 1 pour les individus vivant avec leur famille nucléaire au moment de l'enquête. Le revenu des parents est exprimé en milliers de dollars de 2000.
4. Les résultats sont basés sur l'échantillon pondéré. Tous les scores sont mesurés en rang centile.
5. Les statistiques F sont obtenues pour l'hypothèse nulle d'égalité des effets du score PISA-M et GPA-M ainsi que l'égalité des effets du score PISA-L et GPA-L.
6. Les estimés avec « ** » (*) sont significatifs à 1 % (5 %).

Deuxièmement, nous utilisons les résultats de la régression liant le score PISA et les antécédents au niveau d'éducation atteint afin d'obtenir deux mesures différentes : l'effet d'une variation de 23 centiles dans le score PISA sur le nombre d'années d'études prédit pour un niveau d'éducation parental donné, et la différence de niveau d'études prédit entre les jeunes provenant de familles sans diplôme du secondaire et ceux dont les parents ont un diplôme universitaire, pour un niveau de score PISA fixe.

À partir des effets mesurés et présentés dans la première colonne du tableau 5, on peut conclure qu'une différence de 23 centiles dans le volet mathématiques du test PISA se traduirait par une augmentation de 0,4 année de scolarité. Cet effet est beaucoup plus petit que la différence d'années d'études prédite entre ceux dont les parents sont sans diplôme et ceux dont les parents ont un diplôme d'études supérieures, qui est approximativement égale à une année d'études, puisque l'effet marginal associé à l'indicateur de deuxième cycle est égal à 1,04 et celui du premier cycle est de 0,95.

À l'inverse, il est facile de mesurer combien de centiles de différence au score PISA il faudrait pour générer une année d'études supplémentaire. Les paramètres estimés indiquent qu'il faudrait approximativement un avantage de 50 centiles (la différence entre l'étudiant médian et un étudiant appartenant au premier centile) pour qu'un individu provenant d'une famille sans diplôme obtienne une année d'études supplémentaire (toutes choses étant égales par ailleurs).

Il est aussi pertinent de comparer l'effet du score PISA à celui du revenu familial. L'effet marginal estimé pour chaque millier de dollars est de 0,003 et est précisément estimé. Bien que positif, l'effet du revenu familial semble très faible. Une augmentation de revenu de 100 000 \$ (plus de deux fois l'écart-type de la distribution) n'augmenterait l'éducation que de 0,3 année d'études, et il faudrait donc une différence de revenu de 333 000 \$ pour générer une année d'études supplémentaire⁶.

Les résultats obtenus en analysant les compétences en lecture (colonnes 4, 5 et 6) sont presque identiques à ceux obtenus avec les compétences en mathématiques. Bien que l'effet d'un centile du score PISA en lecture (0,016) soit légèrement supérieur à l'effet mesuré du résultat scolaire (0,013), leurs effets séparés (0,020 et 0,019) sont virtuellement identiques. Tout comme c'était le cas avec les compétences en mathématiques, on peut conclure que ces deux mesures ont un même impact sur le nombre d'années d'études puisque le test statistique pour l'hypothèse d'égalité (0,59 avec une valeur P de 0,44) indique aussi l'équivalence des deux mesures.

Comme auparavant, une différence de 23 centiles augmenterait le nombre d'années d'études d'environ 0,45, et, comme avec la compétence en mathématiques, cet effet est bien inférieur à celui généré par le différentiel parental diplôme supérieur vs sans diplôme, qui se situe lui-même entre 0,8 et 1,2 année⁷.

L'analyse précédente s'est concentrée sur une comparaison entre score PISA et résultats scolaires. Qu'en est-il du facteur non cognitif? Pour comparer l'importance du non cognitif, il est suffisant de se baser sur les estimations des colonnes 1 et 4, c'est-à-dire les spécifications où le score PISA et le GPA sont inclus, puisque les conclusions sont identiques dans les autres spécifications.

⁶ Belzil et Hansen (2021) trouvent aussi des effets très faibles du revenu familial sur l'éducation aux États-Unis dans la cohorte du NLSY97 (à peu près contemporaine de celle étudiée dans l'EJET).

⁷ En combinant les deux scores PISA et en les confrontant aux résultats scolaires globaux (toutes matières confondues), les résultats sont identiques. Leurs effets estimés (0,024 pour PISA et 0,026 pour le GPA total) indiquent le même niveau de similarité et les mêmes capacités prédictives, ce qui mène aussi à l'acceptation de l'hypothèse nulle.

On note que l'effet du facteur non cognitif sur l'éducation est toujours significativement différent de 0 à un niveau de significativité de 1 %. Les effets mesurés, de l'ordre de 0,007 et 0,008, indiquent un effet de rang important, bien que généralement inférieur à celui mesuré pour le score PISA et les résultats scolaires (GPA). Cela est vrai pour l'analyse des compétences en mathématiques et en lecture. Ces effets mesurés représentent généralement entre 50 % et 75 % de l'effet du score PISA ou de l'effet du résultat scolaire. Ces résultats démontrent donc encore l'importance du niveau de motivation et de confiance des jeunes adolescents.

Bien que l'analyse de l'impact des traits non cognitifs ne soit pas l'objectif primordial de notre étude, nous aimerions toutefois faire remarquer que les mesures incluses dans l'EJET ne sont pas nécessairement les plus déterminantes du point de vue de l'économiste. Comme cela a été mentionné plus tôt, le trait non cognitif le plus important est sans doute celui mesurant la motivation personnelle au succès (connu sous le vocable de « conscienciosité » en psychologie de la personnalité), mais qui, malheureusement, n'est pas directement mesuré dans l'EJET. Pour cette raison, l'effet mesuré du facteur non cognitif peut sans doute être considéré comme une borne inférieure de l'importance des traits de personnalité.

Antécédents familiaux et robustesse

L'analyse présentée ci-dessus était basée sur la séparation de l'effet des antécédents familiaux et de l'effet des compétences individuelles et du facteur non cognitif. Comme les mesures cognitives et le facteur non cognitif sont fortement corrélés avec certaines variables, telles que le niveau d'éducation des parents, nous examinons maintenant la sensibilité des résultats à l'exclusion de l'ensemble de ces variables. De plus, nous évaluons aussi si les effets des scores PISA et des résultats scolaires sont aussi sensibles à l'exclusion du facteur non cognitif. Cette question est pertinente dans la mesure où très peu de bases de données individuelles permettent d'isoler le rôle joué par la personnalité des individus.

Pour ce faire, nous reproduisons les colonnes 1 (pour les compétences en mathématiques) et 4 (pour les compétences en lecture) du tableau 5, en excluant toutes les caractéristiques individuelles et parentales. Ces résultats se retrouvent dans les colonnes 1 et 3 du tableau 6. Les estimations se retrouvant dans les colonnes 2 et 4 du tableau 6 sont celles obtenues lorsque le facteur non cognitif est exclu de la régression.

Tableau 6 – Décomposition du niveau d'éducation atteint par la méthode des moindres carrés sans antécédents familiaux

Var. expliquée	Années d'éducation			
	(1)	(2)	(3)	(4)
PISA-M (Math)	0,013**	0,014**	-	-
PISA-L (Lecture)	-	-	0,019**	0,020**
GPA-M (Math)	0,017**	0,020**	-	-
GPA-L (Lecture)	-	-	0,019**	-
Non cognitif	0,006**	-	0,005**	0,021**
Constante	12,686**	12,852**	12,436**	12,534**
R carré	0,16	0,16	0,22	0,22
Nbre d'observations	1752	1752	1763	1763
Stat. F	0,67	1,91	0,04	0,03
Valeur P	0,41	0,17	0,85	0,86

Remarques :

1. Les résultats sont basés sur des régressions obtenues à l'aide de l'échantillon pondéré.
2. Le niveau d'éducation atteint est mesuré en nombre d'années d'études complétées.
3. Les variables P. Univ_2ièmeC, P. Univ_1erC, P. Collégial et P-Secondaire sont des indicateurs binaires du niveau d'éducation des parents. La variable « Famille nucléaire » est égale à 1 pour les individus vivant avec leur famille nucléaire au moment de l'enquête. Le revenu des parents est exprimé en milliers de dollars de 2000.
4. Les résultats sont basés sur l'échantillon pondéré. Tous les scores sont mesurés en rang centile.
5. Les statistiques F sont obtenues pour l'hypothèse nulle d'égalité des effets du score PISA-M et GPA-M ainsi que l'égalité des effets du score PISA-L et GPA-L.
6. Les estimés avec « ** » (*) sont significatifs à 1 % (5 %).

De façon générale, les estimations des colonnes 1 et 3 indiquent que seuls les effets des résultats scolaires sont majorés lorsque les antécédents familiaux sont exclus. C'est le cas à la fois pour les compétences en mathématiques et en lecture. Pour les mathématiques, l'effet du score PISA reporté à la première colonne (0,013) est pratiquement identique à celui du tableau 5 (0,014). En revanche, l'impact de la variable GPA-M est majoré significativement puisque le paramètre augmente d'environ 60 %, passant de 0,011 à 0,017. Cette variation est explicable par la plus grande importance des antécédents familiaux dans les résultats scolaires que dans les scores PISA. Pour les capacités en lecture, l'exclusion des antécédents magnifie légèrement l'effet du score PISA en lecture de 0,016 à 0,019 et augmente un peu plus l'effet des résultats scolaires, qui passe de 0,013 à 0,019.

Cependant, que ce soit pour les compétences en mathématiques ou en lecture, l'exclusion des variables mesurant les antécédents familiaux ne change nullement les résultats des tests statistiques d'égalité des effets des mesures PISA et GPA.

Finalement, la comparaison des résultats des colonnes 2 et 4 avec ceux des colonnes 1 et 3 nous indique qu'une petite partie seulement des effets mesurés pour PISA et GPA est expliquée par le facteur non cognitif. En effet, dans les deux cas (mathématiques et lecture), l'exclusion du facteur non cognitif fait mécaniquement augmenter les effets des variables PISA et GPA, mais ces

augmentations demeurent très petites, en particulier pour les scores PISA, dont les effets mesurés n'augmentent que de 0,001 pour les mathématiques et la lecture. Les conclusions tirées à partir des résultats se trouvant dans le tableau 5 ne sont donc pas vraiment affectées par le choix de modéliser le nombre d'années d'études avec ou sans les antécédents familiaux.

À ce stade, il est intéressant de procéder à des comparaisons de nos résultats avec ceux obtenus pour les États-Unis. Une revue de la littérature utilisant les mesures de compétences standardisées *Armed Services Vocational Aptitude Battery* (ASVAB), desquelles est tiré le célèbre *Armed Forces Qualification Test* (AFQT), indique généralement un rôle plus important du facteur cognitif sur le niveau d'éducation que celui que nous trouvons pour le score PISA (Cameron & Heckman, 1998; Belzil & Hansen, 2021). Cela est peut-être compatible avec l'existence d'une différence de nature entre le test PISA et les items du ASVAB dont le célèbre score AFQT est issu, puisque, comme mentionné précédemment, l'OCDE désire évaluer les capacités individuelles à appliquer des concepts mathématiques dans un contexte pratique, alors que les composantes du ASVAB utilisées pour construire le score AFQT mesurent les performances dans des tâches plus scolaires. Il se peut donc que ces deux tests ne mesurent pas exactement les mêmes facteurs ou ne les mesurent pas dans les mêmes proportions.

Somme toute, notre analyse a révélé un troisième résultat important :

Résultat 3 : Les scores individuels PISA et les résultats scolaires à 15 ans ont des effets limités, mais comparables sur le nombre d'années d'études complétées, en particulier lorsqu'on les compare à l'impact du niveau d'éducation des parents. Une augmentation de 50 centiles permettrait à un individu provenant d'une famille sans diplôme d'obtenir une année d'études supplémentaire (toutes choses étant égales par ailleurs).

8- Le test PISA est-il prédictif du choix de filière?

Pour évaluer la capacité prédictive des scores PISA et des résultats scolaires sur les choix de filières, et tout particulièrement sur la probabilité de diplomation dans une filière scientifique (STIM), nous adoptons deux approches différentes. Dans la première, les effets sont estimés dans la population totale. Dans la seconde, ces mêmes effets sont estimés pour la sous-population des diplômés universitaires. Cette dernière approche est typiquement utilisée dans la littérature microéconométrique sur les choix de filières, qui analyse les décisions individuelles conditionnellement au choix de participation aux études universitaires. Pour les fins du présent rapport, nous n'avons aucune raison de favoriser l'une ou l'autre.

La population totale

Afin d'analyser l'impact des scores PISA dans la population totale, il est naturel de modéliser le choix de façon multinomial. Pour ce faire, nous utilisons un logit multinomial classique avec trois modalités : diplomation universitaire STIM, diplomation universitaire dans toute autre filière (non-STIM) et une modalité alternative englobant toute personne sans diplôme universitaire.

Bien que la modalité alternative puisse être décomposée entre différents types de diplômes, nous ne reproduisons que ceux obtenus avec trois modalités seulement, puisque notre intérêt principal se porte sur la diplomation STIM et que les conclusions demeurent les mêmes avec un plus grand nombre de modalités.

Pour la présentation de nos résultats, nous suivons l'approche standard qui consiste à ne rapporter que les effets marginaux, puisque la non-linéarité du modèle logit multinomial réduit considérablement l'information contenue dans les paramètres estimés. Les résultats obtenus pour la population totale se trouvent dans le tableau 7. Les valeurs présentées s'interprètent donc comme suit. Pour prendre un exemple, un effet marginal de 0,01 impliquerait que la différence moyenne de probabilité (proportion) de diplomation STIM imputable à une différence de rang d'une unité (un centile dans notre modèle) serait de 0,01 (ou une augmentation de 1 %).

Puisque le modèle logit multinomial nous permet d'estimer les effets des caractéristiques individuelles sur les probabilités de choix des trois modalités, nous incluons les effets des compétences en mathématiques sur la probabilité de diplomation STIM dans les colonnes 1, 2 et 3, ainsi que les effets des compétences en lecture sur la probabilité de diplomation universitaire dans toute autre discipline (non-STIM) dans les colonnes 4, 5 et 6.

De plus, et bien que nous ayons utilisé une spécification qui incorpore aussi toutes les caractéristiques individuelles, le tableau 7 n'inclut que les effets des mesures de compétences et l'indicateur du genre (fille), puisque les effets des antécédents familiaux sur les probabilités de choisir une filière donnée sont à peu près tous très faibles et non significatifs⁸.

⁸ Les paramètres mesurant les différences géographiques montrent que les Québécois ont une probabilité plus faible que les Ontariens d'obtenir un diplôme STIM (malgré le fait qu'ils dominent les Ontariens au test PISA_M), mais aussi que cette différence est imprécise.

Tableau 7 – Modélisation du choix de filière avec un modèle logit multinomial : échantillon de population totale

Variable expliquée	Effets marginaux					
	Prob. diplôme STIM			Prob. diplôme non-STIM		
	(1)	(2)	(3)	(4)	(5)	(6)
PISA-Math	0,001*	0,001**	-	-	-	-
PISA-Lecture	-	-	-	0,002**	0,003**	-
GPA-Math	0,001**	-	0,001**	-	-	-
GPA-Lecture	-	-	-	0,002**	-	0,003**
Non cognitif	-0,002	0,000	-0,000	-0,000	-0,000	0,000
Filles	-0,073**	-0,064**	-0,075**	0,149**	0,173**	0,161**
Nbre d'observations	1752	1857	1754	1763	1857	1765

Remarques :

1. Les résultats sont basés sur l'échantillon pondéré. Tous les scores sont mesurés en rang centile.
2. Paramètres du modèle logistique multinomial estimés par la méthode de vraisemblance.
3. Les probabilités dépendent aussi des indicateurs d'éducation des parents, de l'indicateur « Famille nucléaire », du genre et des indicateurs régionaux. La plupart de leurs effets marginaux étant très peu significatifs, nous ne les présentons pas dans le tableau.
4. Les estimés avec « ** » (*) sont significatifs à 1 % (5 %).

Les effets marginaux des colonnes 1 à 3 nous indiquent que les effets associés au score PISA et aux résultats scolaires en mathématiques sont identiques quand l'un ou l'autre est exclu de la spécification. L'effet marginal moyen du score PISA, tout comme celui du GPA-Math, est égal à 0,001. Dans tous les cas, ces effets marginaux sont très précisément estimés. Pour traduire concrètement l'importance de ces effets, il est intéressant de voir leur correspondance associée à un différentiel de rang élevé, tel que celui utilisé dans la section précédente.

Précisément, une différence de rang de 50 centiles (c'est-à-dire la différence de score PISA-Math entre l'élève médian et celui appartenant au premier centile) induirait une augmentation de 5 points de pourcentage dans la probabilité d'obtenir un diplôme universitaire scientifique. Ce qui peut sembler plus étonnant est que l'effet associé au rang centile du GPA-Math n'est pas supérieur, puisqu'il implique la même différence entre l'élève médian et celui du premier centile.

De fait, l'importance relative de ces effets peut être illustrée à l'aide d'une comparaison garçon-fille. Les estimations des colonnes 1, 2 et 3 indiquent qu'à compétence égale en mathématiques, les filles ont une probabilité de diplomation STIM inférieure d'environ 0,07, c'est-à-dire une différence légèrement plus importante que la différence moyenne entre l'étudiant médian et le meilleur.

Il est possible d'interpréter ce résultat de trois façons, qui ne sont d'ailleurs pas mutuellement exclusives. Premièrement, ces effets relativement faibles constituent une indication de plus que les choix éducatifs sont largement déterminés par des aspects autres que purement cognitifs, tels que les goûts et les préférences. Cela est bien documenté pour les États-Unis dans Arcidiacono et coll.

(2012) et Wiswall et Zafar (2016), de même que pour le Canada et le Québec dans Belzil, Hansen et Pernaudet (2025).

Une deuxième explication réside dans le fait que les résultats scolaires obtenus à 14-15 ans sont non standardisés et peuvent donc être suffisamment bruités de façon à réduire leur effet mesuré. Finalement, une troisième interprétation possible est que le curriculum suivi en mathématiques ou le système de notation prévalant dans la première phase du secondaire ne sont pas suffisamment sélectifs pour identifier les élèves qui pourront subséquemment intégrer des études scientifiques.

Les résultats obtenus pour les effets des capacités en lecture sur la probabilité de diplomation dans les filières autres que scientifiques montrent un niveau prédictif plus grand que les capacités en mathématiques. Comme c'est le cas avec les compétences en mathématiques, on ne retrouve aucune différence entre les effets du score PISA et ceux des résultats scolaires en lecture. En revanche, le niveau des effets marginaux obtenus lorsque chacune de ces mesures est obtenue en excluant l'autre, égal à 0,003 dans les deux cas, indique qu'un score élevé au test PISA_L ou un résultat scolaire (GPA_L) plus élevé a un effet trois fois plus élevé sur la diplomation hors sciences que les capacités en mathématiques sur la diplomation STIM.

Pour reprendre l'exemple utilisé pour les mathématiques et impliquant une comparaison entre l'élève médian et celui appartenant au premier centile, une différence de rang de 50 centiles induirait une augmentation de 0,15 dans la probabilité d'obtenir un diplôme universitaire non scientifique. Il est intéressant de noter que ce niveau de différentiel de probabilité de diplomation attribuable à un écart de 50 centiles correspond presque exactement à celui imputable à la différence entre garçon et fille. En effet, et à compétences égales, les filles ont une probabilité d'obtention de diplôme dans une filière autre que scientifique nettement supérieure à celle des garçons. Cette différence est de l'ordre de 0,149 (colonne 1) et de 0,172 (colonne 2).

Sous-population des diplômés

Dans cette sous-section, nous estimons maintenant l'effet des compétences en mathématiques et des compétences verbales sur la probabilité d'obtenir un diplôme scientifique parmi l'ensemble des diplômés universitaires. Ces effets doivent donc être interprétés différemment de ceux décrits ci-dessus, puisqu'ils constituent une analogie de forme réduite de la décision de choix de filière pour ceux qui ont déjà décidé d'entreprendre des études universitaires. Pour cette raison, il est très difficile de prévoir à priori si les effets marginaux des compétences en mathématiques devraient être plus ou moins importants que ceux mesurés dans la population totale, puisqu'ils représentent des paramètres économiques différents.

Puisque notre intérêt principal réside dans la probabilité de diplomation STIM, et pour minimiser les hypothèses paramétriques, nous utilisons un modèle linéaire de probabilité dans lequel la variable dépendante est une variable binaire égale à 1 pour les diplômés STIM et à 0 pour les autres. Le modèle est estimé à partir de l'échantillon d'individus ayant obtenu un diplôme universitaire. Les résultats se trouvent dans le tableau 8, qui a d'ailleurs la même structure que le tableau 7.

Tableau 8 – Modélisation de la probabilité d'obtention d'un diplôme scientifique avec modèle de probabilité linéaire : échantillon des diplômés

Var. dépendante	Diplôme STIM					
	(1)	(2)	(3)	(4)	(5)	(6)
PISA-M (Math)	0,001	0,001*	-	-	-	-
PISA-L (Lecture)	-	-	-	0,001	0,000	-
GPA-M (Math)	0,002**	-	0,002**	-	-	-
GPA-L (Lecture)	-	-	-	-0,001**	-	-0,001*
NC (non cognitif)	-0,001	0,000	-0,001	0,001	0,000	0,001
Filles	-0,195**	-0,183**	-0,198**	-0,172**	-0,188**	-0,167**
Constante	0,267**	0,251**	0,291**	0,291**	0,302**	0,314**
R carré	0,11	0,10	0,11	0,10	0,09	0,09
Nbre d'observations	831	870	832	833	871	833
Stat F	0,73	-	-	3,46	-	-
Valeur P	0,39	-	-	0,05	-	-

Remarques :

1. Les résultats sont basés sur l'échantillon pondéré. Tous les scores sont mesurés en rang centile.
2. Paramètres estimés à l'aide d'un modèle de probabilité linéaire par la méthode des moindres carrés.
3. Les estimés avec « ** » (*) sont significatifs à 1 % (5 %).

De façon générale, on peut affirmer que, parmi la sous-population des diplômés universitaires, les résultats individuels au volet mathématiques du test PISA ont un effet prédictif plus faible sur la fréquentation d'une filière scientifique que les résultats scolaires. En fait, le score du volet mathématiques de PISA ne donne aucune information sur la probabilité de choisir la voie scientifique lorsqu'il est estimé en même temps que l'effet du GPA-M puisque, lorsque les deux mesures de compétences en mathématiques sont incluses dans le modèle (1^{re} colonne du tableau 8), l'effet du score PISA (égal 0,0006, mais arrondi à 0,001 dans le tableau) est non seulement minuscule, mais aussi très imprécis, et l'hypothèse de nullité de l'effet du score au volet mathématiques du test PISA ne peut être rejetée.

Contrairement à la modélisation du niveau d'année d'études et à celle du choix de filière dans la population, l'effet des résultats scolaires en mathématiques est plus marqué que le score PISA. Son effet estimé (de l'ordre de 0,002) est statistiquement significatif, mais demeure faible en niveau. Une différence de 50 centiles se traduirait par une augmentation de la probabilité d'obtention d'un diplôme STIM de 10 points de pourcentage, c'est-à-dire le double de l'effet attribué au score PISA_M.

Cette différence de proportion de 0,10 représente la moitié de la différence garçon-fille, puisque cette dernière est égale à -0,20 pour les filles. Comme on peut le voir au tableau 8, l'imprécision de l'effet du score PISA implique aussi qu'on ne peut rejeter l'hypothèse nulle d'égalité des deux effets.

Les estimations des colonnes 2 (pour le score PISA-Math) et 3 (pour le GPA-Math) confirment la plus grande importance des résultats scolaires, puisque l'effet du GPA en mathématiques (0,002) est exactement le double de l'effet du score PISA (0,001), qui atteint un niveau de significativité pratiquement égal au niveau standard de 5 %.

Ces estimés impliquent donc que, pour les résultats scolaires en mathématiques, un différentiel de 50 centiles se traduit par une majoration de la probabilité d'obtention d'un diplôme STIM de 0,10, tandis que, pour le score PISA, la majoration serait d'environ 0,05. En d'autres termes, dans la sous-population des diplômés universitaires, les résultats scolaires à 14-15 ans semblent deux fois plus importants que le score PISA.

En ce qui concerne les compétences en lecture (colonnes 4, 5 et 6), on observe aussi des effets plus marqués pour les résultats scolaires que pour le score PISA. Le score PISA n'a pas d'effet significativement différent de 0, qu'il soit inclus avec les résultats scolaires (colonne 4) ou non (colonne 5). On peut donc presque dire que la compétence en lecture mesurée par PISA n'est pas moins prédictive de l'obtention d'un diplôme scientifique que la compétence en mathématiques. Cependant, il existe une différence marquée entre les compétences en lecture et celles en mathématiques. Les estimations des colonnes 4 et 6 nous indiquent clairement qu'un meilleur rang dans la performance scolaire en lecture réduit significativement la probabilité de fréquentation d'un programme scientifique de l'ordre de 0,001. Ce résultat est probablement explicable par le fait que les résultats scolaires mesurés à 14 ans sont déjà partiellement marqués par une dimension de goût et un désir de spécialisation qui mènent ultimement à une filière non scientifique.

D'ailleurs, le cas de la compétence en lecture est le seul où l'égalité des effets est très proche de la région de réjection. En effet, la valeur du test F, égale à 3,46, est pratiquement à la frontière du rejet avec une valeur P égale à 0,06⁹.

Pour synthétiser cette section, notre analyse révèle deux résultats importants :

Résultat 4 : Dans la population totale, le score PISA en mathématiques est très peu informatif sur la probabilité d'obtenir un diplôme scientifique et a une capacité prédictive à peu près identique à celle des résultats scolaires. Nos résultats indiquent que l'étudiant appartenant au premier centile des scores PISA n'aurait qu'une probabilité de diplomation STIM supérieure de 0,05 à celle de l'étudiant médian dans la population.

Résultat 5 : Dans la sous-population des diplômés universitaires, les scores PISA-Math sont encore moins discriminants que dans la population totale. Les résultats scolaires en mathématiques sont d'ailleurs deux fois plus importants (prédictifs) que le score PISA.

⁹ Comme c'était le cas dans la population totale, les jeunes Québécois ont une plus faible probabilité d'obtenir un diplôme scientifique que les Ontariens lorsqu'on restreint notre échantillon aux seuls diplômés universitaires. Cependant, ce différentiel est non significatif statistiquement à n'importe quel niveau de confiance réaliste.

9- Le test PISA est-il prédictif du niveau de revenu?

Après avoir analysé l'impact des mesures de compétences mesurées par PISA et celles mesurées en classe sur le succès scolaire tels que le niveau d'études ou la diplomation STIM, nous analysons maintenant les propriétés prédictives de ces mêmes mesures sur les revenus de travail à l'âge de 30 ans. Pour ce faire, nous suivons le même processus que celui adopté pour l'éducation et le choix de filière, c'est-à-dire une comparaison à l'aide d'une régression incorporant les deux mesures et deux autres incorporant chaque mesure séparément. Nos régressions comprennent des contrôles géographiques ainsi qu'un indicateur binaire pour les filles, mais n'incluent pas les indicateurs de niveau d'éducation des parents ni celui de composition familiale (une approche standard dans la littérature).

Il est à noter que nous utilisons une modélisation de forme réduite dans laquelle les revenus ne dépendent que des mesures de compétences et des antécédents mesurés à 15 ans. Une analyse séparant l'effet des compétences de celui des choix éducatifs se trouve dans la prochaine section.

Tableau 9 – Décomposition du log salaire (en 2015) : forme réduite

Var. dépendante	Log salaire (en 2015)					
	(1)	(2)	(3)	(4)	(5)	(6)
PISA-M (Math)	0,003**	0,004**	-	-	-	-
PISA-L (Lecture)	-	-	-	0,003**	0,003**	-
GPA-M (Math)	0,001	-	0,003**	-	-	-
GPA-L (Lecture)	-	-	-	0,001	-	0,002
NC (non cognitif)	0,003**	0,003**	0,003**	0,003**	0,003**	0,004**
Filles	-0,289**	-0,288**	-0,301**	-0,328**	-0,324	-0,316**
Ouest	0,232**	0,235**	0,219**	0,219**	0,210**	0,229**
Ontario	0,318**	0,311**	0,281**	0,279**	0,281**	0,267**
Atlantique	0,273**	0,260**	0,224**	0,238**	0,221**	0,218**
Québec	-	-	-	-	-	-
Constante	10,174**	10,210**	10,305**	10,264**	10,301**	10,333**
R carré	0,09	0,09	0,08	0,08	0,08	0,07
N ^{bre} d'observations	1609	1704	1610	1620	1704	1621
Stat. F	1,13	-	-	1,07	-	-
Valeur P	0,29	-	-	0,30	-	-

Remarques :

1. Les résultats sont basés sur l'échantillon pondéré. Tous les scores sont mesurés en rang centile.
2. Paramètres estimés par la méthode des moindres carrés,
3. Les estimés avec « ** » (*) sont significatifs à 1 % (5 %).

L'analyse des trois premières colonnes indique une différence intéressante avec l'analyse des choix éducatifs pour ce qui est de l'impact du test PISA. En effet, alors que le score PISA-M n'avait pratiquement aucun impact sur le nombre d'années d'études et un impact plus faible que

celui des résultats scolaires en mathématiques sur la probabilité d'obtention d'un diplôme scientifique, son impact sur les salaires est supérieur à celui des résultats scolaires en mathématiques lorsque les deux mesures sont incluses (colonne 1) ou non (colonne 2).

La valeur du paramètre obtenu à la colonne 2 (de l'ordre de 0,004) indique que, toutes choses étant égales par ailleurs, une différence de 50 centiles se traduit par une majoration du revenu de l'ordre de 20 %. Une différence de 35 centiles (équivalente à une différence générée par la variation d'un écart-type pour une loi normale) se traduit par une augmentation de revenu de 14 %.

Lorsque les deux mesures sont incluses, l'effet de rang du GPA_M est trois fois plus petit (0,001) et est non significativement différent de 0. Lorsque le score PISA_M est exclu (colonne 3), l'estimation atteint 0,003 et devient significative, mais demeure inférieure à celle du score PISA.

On retrouve un phénomène encore plus marqué avec les compétences en lecture, puisque le rang individuel au GPA-L n'a aucun effet significatif. L'effet du score PISA-L (0,003), bien qu'inférieur à celui du volet mathématiques, est estimé précisément en colonnes 1 et 2. Son effet associé à une différence de 50 (35) centiles serait équivalent à une augmentation de revenu d'environ 15 % (10 %).

D'ailleurs, l'imprécision du paramètre mesurant l'effet du résultat scolaire en lecture sur les salaires explique en grande partie le faible niveau de la statistique F (égale à 1,07 avec une valeur P de 0,30), qui indique de nouveau qu'on ne peut rejeter l'hypothèse d'égalité.

L'analyse des revenus à l'âge de 30 ans révèle un autre élément marquant. Contrairement à ce qui advenait dans l'analyse des choix éducatifs, les traits non cognitifs semblent jouer un rôle beaucoup plus déterminant dans le succès individuel mesuré par les revenus de travail réalisés pendant le cycle de vie professionnel.

En effet, l'effet d'un centile de trait non cognitif est systématiquement égal à 0,003 sur le log salaire, que ce soit dans le cas où il est conjugué avec l'effet de la compétence en mathématiques (colonnes 1 à 3) ou avec l'effet de la compétence en lecture (colonnes 4 à 6). Cela indique donc que la compétence en mathématiques, bien qu'elle soit plus valorisée que la compétence en lecture, n'est pas plus importante que les traits non cognitifs, tels que la motivation ou la conscienciosité. Il est intéressant de noter que, dans la littérature sur la psychologie des traits de personnalité, il est souvent rapporté que, contrairement à l'intelligence générale qui évolue très peu après 14-15 ans, plusieurs caractéristiques non cognitives, telles que la motivation et la conscienciosité, continuent de se développer. L'importance statistique des traits non cognitifs mesurés dans l'EJET à 14-15 ans pour le succès professionnel qui est mesuré ultérieurement est probablement une borne inférieure puisque, bien que le niveau de motivation à l'âge adulte soit probablement corrélé aux mesures disponibles dans l'EJET, ces dernières ne constituent sans doute qu'une mesure particulièrement bruitée des traits de caractère pertinents au moment de l'entrée dans la vie professionnelle.

Finalement, la valeur des paramètres associés aux différences régionales, et à compétences égales, les salaires québécois sont de l'ordre de 25 % moindre que ceux des autres régions du Canada, y compris les provinces atlantiques.

Résultat 6 : Alors que les scores PISA (en mathématiques aussi bien qu'en lecture) prédisent plus faiblement le succès scolaire ultérieur que les résultats scolaires à 15 ans, la relation est inversée lorsqu'on considère les revenus de travail à 30 ans, puisque les scores PISA peuvent avoir un effet trois fois plus grand que celui des résultats scolaires.

10- Les scores PISA et le biais d'habileté

Après avoir examiné les liens entre compétences, choix éducatifs et revenus, nous analysons les liens entre compétences et rendements économiques de l'éducation. La notion de « biais d'habileté » (un concept central en économie de l'éducation) nous apprend que l'estimation du rendement de l'éducation par la méthode de régression à partir de données ne contenant pas de mesures cognitives (comme des scores individuels standardisés) aura tendance à surestimer l'effet réel causal, puisque les individus les plus doués ont normalement tendance à être plus éduqués. Ce biais est difficile à évaluer directement puisque très peu de bases de données naturelles permettant d'observer le lien entre éducation et salaires contiennent aussi des mesures provenant de tests standardisés. Cependant, puisque l'EJET contient les scores PISA individuels ainsi que les résultats scolaires, il est possible de quantifier la sensibilité des paramètres mesurant l'effet du nombre d'années d'études et des choix individuels de filières à l'inclusion des mesures de compétences. Dans chacun de ces deux cas, nous estimons d'abord les rendements de l'éducation sans mesure de compétence. Nous les estimons ensuite en incluant les deux mesures de compétences en mathématiques, puis une autre fois en incluant les deux mesures de compétences en lecture. Ces résultats se trouvent au tableau 10.

Les estimations obtenues pour le nombre d'années d'études indiquent un rendement annuel de 14 % lorsqu'aucune mesure cognitive n'est incluse (colonne 1). Comme prédit par le concept de biais d'habileté positive, le rendement réel obtenu après l'inclusion de mesures cognitives est réduit substantiellement. Que ce soit en présence des compétences en mathématiques (colonne 2) ou des compétences en lecture (colonne 3), le rendement réel est réduit à 12 % par année d'études et implique donc qu'environ 15 % du rendement mesuré est explicable par les traits cognitifs des individus. Ces résultats sont comparables avec de nombreux résultats obtenus pour divers pays, y compris les États-Unis (Kaymark, 2009).

Afin d'estimer les rendements associés aux filières d'études, nous utilisons la même classification que dans l'analyse du tableau 7 avec trois modalités : les diplômes STIM, les diplômes universitaires non-STIM et une troisième modalité constituant le groupe de référence et englobant tous les individus n'ayant pas de diplôme universitaire.

Lorsque les mesures cognitives sont omises, les diplômés STIM gagnent en moyenne 53 % de plus que les non-diplômés, tandis que les détenteurs d'un diplôme universitaire dans les autres filières (non-STIM) gagnent 38 % de plus (colonne 4). Comme c'est le cas avec le modèle mesurant les rendements à partir des années d'études, ces estimations sont logiquement affectées par un biais positif. Les paramètres se trouvant dans la colonne 5 indiquent effectivement que la prime STIM réelle obtenue après l'inclusion des mesures de compétences en mathématiques est réduite, puisqu'elle passe de 53 % à 42 %. La prime des diplômés non-STIM est aussi réduite, passant de 38 % à 33 %. Pour les valeurs obtenues après l'introduction des mesures de compétences en lecture, on obtient des résultats similaires, puisque les primes réelles diminuent à 0,48 (STIM) et à 0,34 (non-STIM).

Tableau 10 – Biais d'habileté et rendements de l'éducation : années d'études et diplomation scientifique (STIM)

Var. dépendante	Log salaire (2015)					
	(1)	(2)	(3)	(4)	(5)	(6)
PISA-M (Math)	-	0,002**	-	-	0,004**	-
PISA-L (Lecture)	-	-	0,002**	-	-	0,004**
GPA-M (Math)	-	0,002	-	-	0,001	-
GPA-L (Lecture)	-	-	-0,000	-	-	0,001
NC (non cognitif)	-	0,002**	0,002**	-	0,003**	0,003**
Éduc (années)	0,141**	0,118**	0,122**	-	-	-
Expérience	0,101**	0,106**	0,104**	0,114**	0,121**	0,120**
STIM	-	-		0,530**	0,421**	0,479**
Non-STIM	-	-		0,383**	0,317**	0,335**
Filles	-0,372**	-0,351**	-0,359**	-0,267**	-0,288**	-0,302**
Ouest	0,227**	0,228**	0,244**	0,248**	0,244**	0,257**
Ontario	0,267**	0,313**	0,298**	0,291**	0,342**	0,326**
Atlantique	0,167**	0,237**	0,244**	0,224**	0,293**	0,302**
Québec	-	-	-	-	-	-
Constante	7,813**	7,861**	7,800**	9,481**	9,012**	9,038**
R carré	0,21	0,22	0,21	0,16	0,20	0,19
Nbre d'observations	1726	1584	1595	1726	1584	1595

Remarques :

1. Les résultats sont basés sur l'échantillon pondéré. Tous les scores sont mesurés en rang centile.
2. Paramètres estimés par la méthode des moindres carrés.
3. Les estimés avec « ** » (*) sont significatifs à 1 % (5 %).

La réduction de la prime STIM induite par des mesures de compétences semble plus importante en proportion que celle observée pour les autres filières. Cela indique indirectement que le biais d'habileté est plus important pour les diplômés scientifiques que les autres diplômés, et qu'il est compatible avec de nombreux résultats obtenus pour les États-Unis et pour le Canada, qui indiquent que les étudiants des filières scientifiques ont tendance à dominer les autres étudiants dans tous les domaines spécifiques de compétence (Altonji et coll., 2016). Cet avantage absolu favorisant les étudiants STIM implique donc une plus forte corrélation entre les habiletés individuelles et le choix d'une filière scientifique.

Résultat 7 : Pour la cohorte née en 1984, un diplôme scientifique (STIM) induit une prime de revenu de 54 % par rapport à l'ensemble des travailleurs sans diplôme universitaire, tandis qu'un diplôme obtenu dans une filière non scientifique induit une prime de 38 %. Cependant, le différentiel réel prévalant après la prise en compte des compétences mesurées à 15 ans est de 42 % pour un diplôme STIM et de 33 % pour les autres filières universitaires.

11- Conclusions et implications pour le design du test PISA

Les résultats divulgués dans le cadre du programme PISA de l'OCDE sont souvent au centre de controverses économiques, politiques et même statistiques. La divulgation étant confinée à l'échelle nationale, elle induit tout naturellement des comparaisons internationales. Cependant, en partant du postulat qu'un plus haut résultat PISA à l'échelle nationale est un indicateur d'un plus haut niveau moyen de compétences (capital humain) dans un pays donné, il s'en suit logiquement que les différences individuelles observées à l'intérieur d'un pays devraient aussi être assimilées à des différences de capital humain qui, elles-mêmes, devraient se traduire par différents niveaux de succès économique ou scolaire sur le plan individuel. Cela devrait être le cas, malgré l'imprécision statistique inhérente au score individuel PISA auquel nous devons faire face. De plus, la disponibilité des résultats scolaires contemporains aux scores PISA nous a permis d'utiliser ces derniers comme une sorte de référence (*benchmark*) dans nos comparaisons.

L'une des critiques du programme PISA les plus fondées scientifiquement (statistiquement) est que les scores individuels agrégés pour fin de comparaison internationale mesurent l'intelligence fondamentale beaucoup plus que les habiletés spécifiques et que, par conséquent, il est difficile de les utiliser pour inférer (et comparer) la qualité de différents systèmes éducatifs dans des domaines spécifiques.

Dans ce rapport, nous avons tenté de suivre une approche essentiellement économique en répondant aux deux questions fondamentales suivantes :

Les résultats individuels au test PISA mesurés à 15 ans sont-ils capables de prédire les succès scolaires et professionnels ultérieurs ?

Dans quelle mesure les scores individuels PISA donnent-ils de l'information différente de celle déjà contenue dans les résultats scolaires contemporains à l'administration du test PISA et dans les mesures de traits non-cognitifs ?

Les résultats fondamentaux de notre étude peuvent se synthétiser comme suit. Premièrement, le score individuel au volet mathématiques du test PISA est modérément informatif du succès scolaire, surtout si on le compare aux résultats scolaires mesurés à 15 ans. Par exemple, il est pratiquement incapable de prédire l'obtention individuelle d'un diplôme universitaire dans une filière scientifique (un diplôme STIM). En fait, bien qu'il soit standardisé, le test PISA est moins informatif que les résultats scolaires, qui sont, eux, potentiellement plus bruités, bien qu'il soit généralement impossible de rejeter l'hypothèse nulle que les scores PISA et les résultats scolaires ont des effets identiques sur les indicateurs de succès ultérieur.

Deuxièmement, nos résultats indiquent que les scores PISA individuels (en mathématiques ou en lecture) sont beaucoup plus susceptibles de prévoir le succès économique sur le marché du travail (vers 30 ans) que les résultats scolaires.

En conclusion, nos résultats portent un message sans doute moins négatif que celui qu'on retrouve dans une certaine littérature en psychologie quantitative (psychométrie). Bien qu'il soit sans doute vrai que les résultats individuels sont explicables partiellement par le niveau d'habileté générale, leurs bonnes performances prédictives du succès sur le marché du travail indiquent que l'OCDE

est tout au moins partiellement justifiée de considérer le projet PISA comme mesurant des capacités individuelles à résoudre des problèmes concrets et pratiques, et non simplement des compétences découlant directement des programmes scolaires. Néanmoins, on ne peut exclure que ces mêmes capacités soient assimilables à une autre facette, moins théorique et plus concrète, de l'intelligence individuelle.

À partir de nos résultats et des nombreuses analyses statistiques des items individuels synthétisées dans les premières sections de ce rapport, il est justifié de formuler les recommandations suivantes.

Premièrement, et afin de dissiper la confusion possible entre les effets de l'intelligence générale et les compétences spécifiques, il serait judicieux d'insérer un certain nombre d'items spécifiquement consacrés à la mesure de l'intelligence générale afin qu'il soit possible de purger les résultats individuels de son effet. Cela pourrait se faire à l'aide de tests déjà existants et internationalement reconnus (tests de Wechsler, des matrices de Raven, de Stanford-Binet, etc.).

Deuxièmement, et avec l'aide de ces items, il serait possible d'identifier les éléments (questions) du test PISA qui sont le moins affectés par l'intelligence générale, c'est-à-dire ceux qui reposeraient moins (qui « chargeraient » moins en langage statistique) sur les mesures de QI. Cela permettrait d'améliorer le design actuel en augmentant le nombre d'items indicatifs des facteurs ciblés spécifiquement par l'OCDE.

Références

- Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. (2011). Chapter 1 - Personality psychology and economics. *Handbook of the economics of education*, 4, 1-181.
- Altonji, J.G., Arcidiacono, P., & Maurel A. (2016). The analysis of field choice in college and graduate school: determinants and wage effects, *Handbook of the Economics of Education*, 5, 305–396.
- Arcidiacono, P., Hotz, V., & Kang S. (2012). *Affirmative action and sorting across colleges: the effects of Prop 209 on graduation rates among University of California campuses*. Unpublished manuscript, Duke University.
- Belzil, C., & Hansen, J. (2021). The evolution of the US family income-schooling relationship and educational selectivity, *Journal of Applied Econometrics*, 35(7), 841-859
- Belzil, C., Hansen, J., & Pernaudet J. (2025). *Separating the structural and composition effects of randomized higher education financial aid generosity on field of study*, Working Paper.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Lawrence Erlbaum Associates Publishers.
- Borghans, L., Golsteyn, B. H. H., Heckman, J. J., & Humphries, J. E. (2011). Identification problems in personality psychology. *Personality and individual differences*, 51(3), 315-320.
- Brunner, M. (2008). No g in education? *Learning and individual differences*, 18(2), 152–165. <https://doi.org/10.1016/j.lindif.2007.08.005>
- Cameron, S., & Heckman, J. (1998). Life cycle schooling and dynamic selection bias: models and evidence for five cohorts of American males, *Journal of Political Economy*, 106(1998), 262-333.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Cromley, J. G. (2009). Reading achievement and science proficiency: international comparisons from the Programme on International Student Assessment. *Reading Psychology*, 30(2), 89-118. <https://doi.org/10.1080/02702710802274903>
- Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78, 883-931.
- Egelund, N. (2008). The value of international comparative studies of achievement – a Danish perspective. *Assessment in education: Principles, policy & practice*, 15(3), 245–251. <https://doi.org/10.1080/09695940802417400>

- Frey, M. C. (2019). What we know, are still getting wrong, and have yet to learn about the relationships among the SAT, intelligence and achievement. *Journal of Intelligence*, 7(4), 26.
- Hansen, K. T., J. J. Heckman, & Mullen, K. J. (2004). The effect of schooling and ability on achievement test scores. *Journal of Econometrics*, 121(1-2), 39-98.
- Heckman J.J., Stixrud J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 411–82.
- Hopfenbeck, T. N., & Kjærnsli, M. (2018). Students' test motivation in PISA: The case of Norway. *The Curriculum Journal*, 27(3), 406–422.
<https://doi.org/10.1080/09585176.2016.1156004>
- Kaymark, B. (2009). Ability bias and the rising education premium in the United States: A cohort-based analysis. *Journal of Human Capital*, University of Chicago Press, 3(3), 224-267.
- Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive g and academic achievement g one and the same g? An exploration on the Woodcock–Johnson and Kaufman tests. *Intelligence*, 40(2), 123–138.
<https://doi.org/10.1016/j.intell.2012.01.009>
- Gneezy, U., List, J. A., Livingston, J.A., Qin, Xiangdong, Sadoff, S., & Xu, Yang. (2019). Measuring success in education: The role of effort on the test itself. *American Economic Review: Insights*, 1(3), 291–308.
- Grek, S. (2009). Governing by numbers: The PISA Project in Europe. *Journal of Education Policy*, 24(1), 23-37.
- Gustafsson, J. E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, 28(4), 407–434.
http://dx.doi.org/10.1207/s15327906mbr2804_2
- Heckman, J., Jagelka, T., & Kautz, T. (2019). Some contributions of economics to the study of personality, *NBER Working Paper* 26459.
- Holzinger, K., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, 2, 41–54.
- Hunt, E. (2010). *Human intelligence*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511781308>
- Koenig, K. A., Frey, M. C., & Detterman, D. K. (2008). ACT and general cognitive ability. *Intelligence*, 36(2), 153–160. <https://doi.org/10.1016/j.intell.2007.03.005>
- Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "'General intelligence,' objectively determined and measured", *Journal of*

Personality and Social Psychology, 86(1), 96-111. <https://psycnet.apa.org/doi/10.1037/0022-3514.86.1.96>

OECD. (2007). *PISA 2006. Science competencies for tomorrow's world (Volume 1)*. Paris: OECD Publishing.

OECD. (2013a). *PISA 2012 Results: Ready to learn: Students engagement, drive and self-beliefs (Volume III)*. Paris: OECD Publishing.

OECD. (2013b). What is PISA? *PISA 2012 results: What students know and can Do – Student performance in mathematics, reading and science (Volume I)*, Paris: OECD Publishing, Chapter 1, 23-30.

OECD Publishing. Organisation for Economic Cooperation and Development. (2001). OECD employment and labour market statistics. Data available from http://www.oecd-ilibrary.org/employment/data/oecd-employment-and-labour-market-statistics_ifs-data-en

Okubo, T. (2022). Theoretical Considerations on Scaling Methodology in PISA, *OECD Education Working Papers*, 282, Paris: OECD Publishing.

Pokropek, A., Marks, G. N., & Borgonovi, F. (2021). How much do students' scores in PISA reflect general intelligence and how much do they reflect specific abilities? *Journal of Educational Psychology*, 114(5), 1121–1135. <https://doi.org/10.1037/edu0000687>.

Pokropek, A., Marks, G. N., Borgonovi, F., Koc, P., & Greiff, S. (2022). General or specific abilities? Evidence from 33 countries participating in the PISA assessments. *Intelligence*, 92, <https://doi.org/10.1016/j.intell.2022.101653>

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: The Danish Institute of Educational Research.

Rindermann, H., & Baumeister, A. E. E. (2015). Validating the interpretations of PISA and TIMSS tasks: A rating study. *International Journal of Testing*, 15(1), 1–22. <https://doi.org/10.1080/15305058.2014.966911>

Rindermann, H. (2007). The g-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-Tests across Nations, *European Journal of Personality*, 21, 667–706.

Schleicher, A. (2007). Can competencies assessed by PISA be considered the fundamental school knowledge 15-year-olds should possess? *Journal of Educational Change*, 8(4), 349–357. <https://doi.org/10.1007/s10833-007-9042-x>

Schmidt, W.H., & Burroughs, N.A. (2016). The trade-off between excellence and equality: What international assessments tell us, *Georgetown Journal of International Affairs*, 17(1) 103-109.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, 15, 201–293.

Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership*, 41(8), 19–27.

Warne R.T., & Burningham, C. (2019). Spearman's g found in 31 non-Western nations: Strong evidence that g is a universal phenomenon. *Psychological Bulletin*, 145(3), 237-272. [doi: 10.1037/bul0000184](https://doi.org/10.1037/bul0000184).

Wiswall, M., & Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, 82(2), 791–824.

Zaboski, B. A., Kranzler, J. H., & Gage, N. A. (2018). Meta-analysis of the relationship between academic achievement and broad abilities of the Cattell-horn-Carroll theory. *Journal of School Psychology*, 71, 42–56. <https://doi.org/10.1016/j.jsp.2018.10.001>