

# Étude de la persévérance des élèves de 15 ans au test de PISA 2012 et 2015 selon le mode d'administration

Christian Monseur and Yannick Lonhay

Volume 44, Number 1, 2021

URI: <https://id.erudit.org/iderudit/1087988ar>

DOI: <https://doi.org/10.7202/1087988ar>

[See table of contents](#)

Publisher(s)

ADMEE-Canada

ISSN

0823-3993 (print)

2368-2000 (digital)

[Explore this journal](#)

Cite this article

Monseur, C. & Lonhay, Y. (2021). Étude de la persévérance des élèves de 15 ans au test de PISA 2012 et 2015 selon le mode d'administration. *Mesure et évaluation en éducation*, 44(1), 29–59. <https://doi.org/10.7202/1087988ar>

Article abstract

*Based on PISA 2012 and 2015 data, this paper analyzes students' perseverance in taking a low-stakes assessment, defined as their ability to maintain a constant rate of correct answers throughout the test. It also investigates the possible effect of the administration mode (paper-and-pencil vs. computerized test) on participants' perseverance. For that purpose, this research exploits the balanced incomplete block design of PISA 2012 and 2015, and focuses on the answers to the questions posed at each of the four positions. The results show an overall decrease in student performance, regardless of the domain and the cycle, and an increase in variability. However, to the questions asked at the start of the test, the students generally answered better in the 2012 than 2015 cycle. On the contrary, to the questions asked at the end of the test, this trend seems to be reduced or even reversed, which could reflect a greater pupils' motivation to answer with a computerized test.*

## Étude de la persévérance des élèves de 15 ans au test de PISA 2012 et 2015 selon le mode d'administration

Christian Monseur

Yannick Lonhay

Université de Liège

MOTS CLÉS : évaluation à faibles enjeux, persévérance, mode d'administration, PISA 2012, PISA 2015

*À partir des données PISA 2012 et 2015, cette étude analyse la persévérance des élèves à répondre à un test à faibles enjeux, c'est-à-dire leur capacité à maintenir un taux constant de réponses correctes tout au long de l'épreuve. Elle étudie également l'éventuel effet du mode d'administration (papier-crayon comparativement à support informatique) sur la persévérance des participants. À cette fin, cette recherche exploite le plan d'évaluation incomplet balancé de PISA 2012 et 2015, et se concentre sur les réponses aux questions posées à chacune des quatre positions. Les résultats révèlent une diminution globale de la performance des élèves, quels que soient le domaine évalué et le cycle, ainsi qu'une augmentation de sa variabilité. Néanmoins, aux questions posées en début d'épreuve, les élèves ont généralement mieux répondu lors du cycle 2012 qu'au cycle 2015. Au contraire, aux questions posées en fin d'épreuve, cette tendance semble se réduire, voire s'inverser, ce qui pourrait traduire une plus grande motivation des élèves à répondre avec une épreuve informatisée.*

KEY WORDS: low-stakes assessment, perseverance, administration mode, PISA 2012, PISA 2015

*Based on PISA 2012 and 2015 data, this paper analyzes students' perseverance in taking a low-stakes assessment, defined as their ability to maintain a constant rate of correct answers throughout the test. It also investigates the possible effect of the administration mode (paper-and-pencil vs. computerized test) on participants' perseverance. For that purpose, this research exploits the balanced incomplete block design of PISA 2012 and 2015, and focuses on the answers to the questions posed at each of the four positions. The results show an overall decrease in student performance, regardless of the domain and the cycle, and an increase in variability. However, to the questions asked at the start of the test, the students generally answered better in the 2012 than 2015 cycle. On the contrary, to the questions asked at the end of the test, this trend seems to be reduced or even reversed, which could reflect a greater pupils' motivation to answer with a computerized test.*

PALAVRAS-CHAVE: avaliação de baixo risco, perseverança, modo de administração, PISA 2012, PISA 2015

*Usando dados do PISA 2012 e 2015, este estudo analisa a perseverança dos alunos em responder a um teste de baixo risco, ou seja, a sua capacidade de manter uma taxa constante de respostas corretas ao longo do teste. Estuda também o possível efeito do modo de administração (papel-lápis em comparação com o suporte informático) na perseverança dos participantes. Para tal, esta investigação explora o plano de avaliação incompleto equilibrado do PISA 2012 e 2015 e concentra-se nas respostas às questões colocadas em cada uma das quatro posições. Os resultados revelam uma diminuição global do desempenho dos alunos, independentemente do domínio avaliado e do ciclo, bem como um aumento da sua variabilidade. No entanto, às questões colocadas no início do teste, os alunos responderam geralmente melhor durante o ciclo de 2012 do que no ciclo de 2015. Pelo contrário, às questões colocadas no fim do teste, esta tendência parece atenuar-se, mesmo inverter-se, o que poderá refletir uma maior motivação dos alunos para responderem a um teste informatizado.*

## Introduction

Depuis 20 ans, le Programme international pour le suivi des acquis des élèves (PISA), mené sous l'égide de l'Organisation de coopération et de développement économiques (OCDE), évalue les élèves de 15 ans afin de déterminer s'ils « possèdent certaines des connaissances et compétences essentielles pour participer pleinement à la vie de nos sociétés modernes » (OCDE, 2016, p. 12).

Plusieurs des indicateurs PISA constituent des informations précieuses pour aider les responsables politiques à piloter leur système éducatif. Toutefois, puisque ces évaluations ne comportent généralement aucun enjeu majeur pour les élèves, certains participants pourraient ne s'y investir que partiellement. Ce manque d'investissement pourrait s'observer particulièrement dans les systèmes éducatifs qui entretiennent la culture des notes scolaires. En effet, comme le soulignent Fumel et Keskskaik (2017) ou Keskskaik et Rocher (2015), dans ces systèmes éducatifs, par exemple celui de la France, puisque les notes scolaires occupent une place importante, « la question de la motivation des élèves face à une évaluation sans enjeux pour eux mérite d'être posée » (Keskskaik et Rocher, 2015, p. 119). Par conséquent, les compétences des répondants pourraient être sous-estimées et la comparabilité internationale pourrait être altérée.

Les résultats d'études relatives à la motivation des élèves lors d'évaluations à faibles enjeux ne convergent malheureusement pas (Asseburg et Frey, 2013 ; Butler et Adams, 2007 ; Chen, 2002 ; Eklöf, 2007 ; Eklöf et Nyroos, 2013 ; Hopfenbeck et Kjærnsli, 2016 ; Wise et DeMars, 2005), probablement en partie à cause de différences méthodologiques ou de la diversité des outils utilisés pour mesurer la motivation ou l'engagement des élèves (Penk et Schipolowski, 2015).

La présente étude se concentre sur la persévérance des élèves, considérée selon certaines théories comme une des composantes de la motivation (p. ex., la théorie de la valeur attendue ou *expectancy-value theory* de Wigfield et Eccles, 2000). Cette mesure est particulièrement intéressante,

car elle repose uniquement sur des informations objectives et non auto-rapportées relatives à l'exactitude des réponses (Debeer et al., 2014). Aucun autre *artefact* ne doit être mis en œuvre pour appréhender le niveau de motivation des répondants et son estimation n'est donc pas affectée par certains biais culturels (p. ex., la tendance à l'acquiescement) généralement observés pour les données auto-rapportées (Heine et al., 2002).

Par ailleurs, l'enquête PISA a également pour objectif de proposer des indicateurs de tendance conçus afin d'observer l'évolution des performances des différents systèmes éducatifs, notamment pour permettre aux responsables politiques d'évaluer les effets de leur réforme. Cette comparaison temporelle nécessite que les méthodologies employées lors des différents cycles ne puissent pas altérer la comparabilité des résultats. Tout changement méthodologique jugé nécessaire pour telle ou telle raison doit être accompagné des preuves empiriques qui démontrent sa neutralité sur les résultats.

Ainsi, en 2015, l'évaluation PISA est passée d'une version dite papier-crayon à une version informatisée. Des analyses menées au départ des données de l'essai sur le terrain n'ont pas permis de détecter, tous pays confondus, un fonctionnement différentiel de la majorité des items ou, plus globalement, de l'épreuve selon le mode d'administration (OECD, 2017).

Toutefois, la numérisation de l'épreuve pourrait accroître l'investissement de certains élèves et le décroître chez d'autres. En d'autres termes, le mode d'administration pourrait interagir avec des variables telles que le pays, le genre ou même le niveau de performance de l'élève. La littérature ne permet pas à ce jour de conclure à l'existence ou non d'un effet du mode d'administration sur la performance des élèves (Kingston, 2008 ; Wang et al., 2007, 2008). Dès lors, le changement de mode d'administration opéré dans PISA représente une opportunité unique pour analyser cet effet sur la persévérance des élèves à partir de données d'échantillons représentatifs de très grande taille.

## Revue de littérature

Selon la théorie de la valeur attendue, la motivation des personnes qui s'engagent dans une tâche dépend du résultat attendu et du niveau de valorisation de l'activité (Wigfield et Eccles, 2000). La valeur perçue d'une activité pour un individu donné dépend notamment de l'importance et de l'utilité

qu'il peut percevoir dans cette activité. Pour des jeunes de 15 ans, les tests PISA ne comportent généralement aucun enjeu puisque, dans la plupart des pays, ils ne reçoivent aucun incitatif, qu'il soit financier ou autre, et leur performance au test n'influencera pas leur parcours scolaire. On peut dès lors suspecter qu'un certain nombre d'élèves attribuent peu de valeur à ce test et s'y engagent peu ou prou. Si chaque élève, quels que soient ses origines ethniques ou sociales ou encore son pays de résidence, percevait la même valeur à ces épreuves internationales, la comparabilité des résultats ne serait pas remise en cause. Par contre, puisque cette valeur et cette motivation varient d'un pays à l'autre, d'une école à l'autre, voire d'un élève à l'autre, d'aucuns peuvent s'inquiéter de la comparabilité des résultats.

Différentes recherches expérimentales ou quasi-expérimentales ont étudié l'impact des enjeux d'un test sur la motivation et/ou la performance des répondants. Les participants sont généralement répartis aléatoirement en différents groupes dits expérimentaux (lors de méthodologies expérimentales) ou par appariement (dans le cadre de recherches quasi-expérimentales) afin de permettre une manipulation de la variable « enjeux ». Dans la majorité des cas, ces études ont révélé que les enjeux associés aux tests pourraient avoir un impact sur la motivation et sur les performances des élèves. Toutefois, sans réelle surprise, cette influence semble varier en fonction (i) de certaines caractéristiques des élèves (Braun et al., 2011 ; Fumel et Keskaik, 2017 ; Steedle et Grochowalski, 2017), (ii) de l'année d'étude des élèves (O'Neil et al., 1996 ; O'Neil et al., 2005), voire (iii) du système éducatif (List et al., 2019). Ce dernier résultat renforce la nécessité de poursuivre la recherche sur la motivation des répondants, particulièrement dans les enquêtes internationales telles que PISA, notamment afin d'éventuellement différencier dans les résultats observés l'effet de la compétence et l'effet de la motivation.

D'autres études (Asseburg et Frey, 2013 ; Eklöf, 2007, 2015 ; Eklöf et Nyroos, 2013 ; Eklöf et al., 2014 ; Hopfenbeck et Kjærnsli, 2016) ont été menées pour tenter de mesurer la motivation des répondants lors d'évaluation à faibles enjeux en employant notamment des mesures auto-rapportées d'effort. Ce type de mesure est généralement collecté à travers un ou plusieurs questionnaires. Bien souvent, ces questionnaires proposent plusieurs items permettant d'évaluer le ou les construits souhaités (p. ex., l'effort déployé pour répondre au test cognitif, l'importance du test aux yeux des participants, etc.).

Dans certains cycles de PISA, un tel questionnaire, dénommé « thermomètre à l'effort », est présenté à la fin du test cognitif et est composé de deux questions construites sous forme d'un différenciateur sémantique d'Osgood allant de 1 à 10. La première question sert à percevoir l'effort que les répondants ont déployé pour réaliser le test cognitif, tandis que la seconde concerne l'effort qu'ils auraient fourni si la note obtenue avait compté pour leur bulletin. De ces études, il ressort tout d'abord que la difficulté des questions pourrait influencer la motivation auto-rapportée des répondants (Asseburg et Frey, 2013), les étudiants les plus faibles rapportant une moins grande motivation, avec une sous-estimation de leur compétence comme conséquence probable pour cette sous-population.

D'autres études ont montré que la motivation et son effet (au sens statistique du terme) sur la performance varient selon le genre de l'élève (Eklöf, 2007 ; Eklöf et Nyroos, 2013 ; Eklöf et al., 2014 ; Hopfenbeck et Kjærnsli, 2016) ou son pays d'origine (Eklöf, 2015 ; Eklöf et al., 2014). Toutefois, ces résultats ne sont pas confirmés par toutes les études.

Comme susmentionné, ces dernières études peuvent souffrir de différents biais. Ainsi, à titre d'exemple, les données auto-rapportées peuvent être entachées de biais culturels. Par ailleurs, les méthodologies choisies pour modéliser la motivation des répondants peuvent également influencer les résultats obtenus. Butler et Adams (2007) ont ainsi exploré la relation entre les efforts auto-rapportés et les performances lors des cycles 2000 et 2003 de PISA. À partir des deux mesures dudit « thermomètre à l'effort », ils dérivent un indice d'effort relatif qu'ils considèrent comme le reflet de l'effort exprimé par les élèves. Cet indice relatif, comparativement à l'indice absolu, varie moins d'un pays à l'autre, ce qui témoigne d'une moins grande sensibilité aux différences culturelles. Par ailleurs, comme on pouvait s'y attendre, les données PISA indiquent que les élèves signalent qu'ils fournissent globalement moins d'effort lors de ce type d'évaluation que dans un test où leurs résultats compteraient pour leur bulletin (c.-à-d. aux enjeux élevés). Toutefois, d'après Dierendonck et al. (2013) et Keskaik et Rocher (2015), cette variable est plutôt représentative de la différence d'effort déclaré en fonction de l'enjeu du test et ne reflète pas le niveau global d'effort des élèves. Ces auteurs donnent comme exemple qu'un effort relatif de 10 ne signifie pas automatiquement que le répondant fournit un effort maximal, mais plutôt qu'il n'y a pas de différence dans l'effort qu'il exprime en fonction de l'enjeu du test. Ainsi, un élève qui rapporte avoir fourni un effort de 1 pour répondre au test PISA et qui

aurait fourni le même type d'effort (donc 1) si le test avait compté pour des points reçoit la note maximale d'effort relatif (c.-à-d. 10), alors qu'il déclare plutôt un effort minimal.

Afin de dépasser ces faiblesses méthodologiques, une autre approche exploite la rotation des items dans les tests tels que PISA et dérive une mesure objective de la persévérance, soit en comparant l'évolution du pourcentage de réponses correctes selon la position des items au sein du test, soit en intégrant directement dans la modélisation selon la théorie de la réponse à l'item (TRI ; en anglais, *item response theory* ou IRT) un paramètre de position de l'item, qui traduit pour chaque élève la réduction de la probabilité de réussite aux items. Ce paramètre peut donc être assimilé à la persévérance de l'élève. Les résultats des études utilisant l'une de ces deux approches (Borgonovi et Biecek, 2016 ; Debeer et al., 2014 ; Nagy et al., 2019) suggèrent que les taux ou probabilité de bonnes réponses diminuent généralement au fur et à mesure du test et que l'ampleur de cette diminution varie premièrement d'un domaine à l'autre, voire d'un type de questions à l'autre et, deuxièmement, d'un élève à l'autre, d'une école à l'autre ou d'un pays à l'autre.

Plus précisément, des différences plus importantes sont observées en lecture, tandis que les différences les plus faibles sont observées en mathématiques (Nagy et al., 2019). Pour Hohensinn et al. (2011), la position des items n'aurait aucun effet significatif dans un test en mathématiques. Par ailleurs, le déclin de la performance est plus prononcé dans les items à réponse construite que pour les questions à choix multiples (Borgonovi et Biecek, 2016).

Concernant le genre des élèves, une plus grande persévérance est observée auprès des filles (Nagy et al., 2019), en particulier dans le domaine de la lecture (Borgonovi et Biecek, 2016), auprès des élèves issus de milieux socio-économiques favorisés (Borgonovi et Biecek, 2016 ; Nagy et al., 2019) et auprès des natifs (Nagy et al., 2019). Dès lors, on ne peut guère s'étonner d'observer une moindre persévérance dans les filières qualifiantes et dans les écoles comptant une forte proportion d'élèves issus de l'immigration (Nagy et al., 2019). Enfin, la persévérance corrèle positivement avec la performance, tant au niveau des élèves (Borgonovi et Biecek, 2016) qu'au niveau des écoles, du moins dans la plupart des pays (Debeer et al., 2014). À l'échelle des pays, la persévérance serait aussi associée positivement à la performance moyenne (Debeer et al., 2014).



D'autres études se sont également penchées sur l'effet du mode d'administration sur la motivation des élèves (Khoshsima et Hashemi, 2017). À ce jour, trois méta-analyses ont été conduites sur cette problématique, mais aucun résultat ne semble se dégager clairement. Ainsi, Kingston (2008), qui synthétise les résultats de 81 études réalisées entre 1997 et 2007, observe des différences significatives de performance en fonction du format de test. En outre, le domaine évalué pourrait modérer cet effet. Plus précisément, les résultats à une épreuve informatisée seraient supérieurs pour les domaines langue anglaise (ampleur de l'effet moyenne de 0,11) et études sociales (0,15). À l'opposé, de meilleures performances sont observées à une épreuve du type papier-crayon pour le domaine des mathématiques (-0,06). Par contre, les méta-analyses de Wang et al. (2007 ; 2008) ne révèlent aucune différence significative, quel que soit le domaine évalué (mathématiques et lecture). Selon Kingston (2008), « il n'est pas surprenant (compte tenu de la diversité des problèmes de mesure et d'échantillonnage pouvant affecter une étude donnée) que les résultats des études n'aient pas toujours été cohérents » (p. 22, trad. libre).

## Présente étude

### *Hypothèses*

Comme susmentionné, la mesure de la motivation d'individus à partir de données auto-rapportées est influencée par différents types de biais qui remettent en question son utilisation pour l'étude de son influence sur la performance. Dans le cadre de cette recherche, l'étude de la motivation se limitera à une de ses composantes, à savoir la persévérance des élèves, définie, toutes choses égales par ailleurs, comme la capacité des élèves à maintenir un taux de réussite constant tout au long d'un test (Borgonovi et Biecek, 2016). Contrairement aux mesures auto-rapportées, la mesure de la persévérance telle qu'elle est proposée repose uniquement sur des informations objectives, à savoir l'exactitude des réponses.

Sur base de la revue de la littérature et des possibilités offertes par le cadre méthodologique des études PISA, quatre hypothèses sont formulées :

**Hypothèse 1 :** À partir des résultats rapportés par Borgonovi et Biecek (2016) et Debeer et al. (2014), nous postulons que la persévérance des élèves décline constamment au fur et à mesure du test, quels que soient le domaine évalué et le cycle d'évaluation. Ainsi, les

performances moyennes diminueraient progressivement de la première position à la dernière position ;

**Hypothèse 2 :** Puisque la persévérance corrèle positivement avec la performance (Borgonovi et Biecek, 2016), la variabilité de la performance devrait s'accroître de la position 1 à la position 4 ;

**Hypothèse 3 :** Tout comme le suggèrent Braun et al. (2011) ainsi que Nagy et al. (2019), nous soutenons que les performances en mathématiques dépendent davantage du cursus scolaire, comparativement à la compréhension de l'écrit. De plus, les items en lecture sont généralement plus longs que les questions de mathématiques, ce qui requiert un investissement plus élevé de la part des élèves. Dès lors, une moindre persévérance devrait être observée en compréhension de l'écrit ;

**Hypothèse 4 :** Dans le prolongement des résultats rapportés par Kingston (2008) relatifs aux effets du mode d'administration, nous postulons une plus grande persévérance aux épreuves informatisées. Les performances moyennes diminueraient donc de façon moins importante lors du cycle 2015 (test électronique) qu'au cycle 2012 (test papier-crayon) de PISA et l'accroissement de la variabilité devrait être de moindre amplitude en 2015.

## Méthodologie

### *Échantillon*

La présente étude exploite les données PISA des cycles de 2012 et de 2015. Seuls les pays de l'OCDE ont été retenus (hormis le Chili et le Mexique, qui ont dû être écartés, car les carnets de test utilisés par ces deux pays étaient partiellement différents [OECD, 2014]). L'échantillon pour cette étude comprend 259 060 élèves pour le cycle 2012 et 233 999 élèves pour le cycle 2015, tous âgés de 15 ans.

### *Méthode et instrument*

Comme susmentionné, généralement, l'estimation de la persévérance se base sur l'analyse des taux de réponses correctes, plus particulièrement sur la variation des taux de réussite observés en début et en fin de test. Depuis 2003, PISA propose un plan d'évaluation incomplet balancé pour les épreuves cognitives. Cette méthode consiste, d'une part, à soumettre un sous-ensemble de questions à chaque élève et, d'autre part, à agencer les blocs de questions

de telle sorte qu'ils apparaissent le même nombre de fois dans chacune des positions. Puisque chaque question apparaît dans chacune des positions, elles sont affectées de manière semblable par l'effet d'ordre. La comparabilité de leur difficulté n'est donc pas altérée et l'effet est en quelque sorte contrôlé.

Ainsi, pour le cycle 2012, le plan d'évaluation comportait 13 livrets de tests, chacun composé de 4 blocs de questions. L'ensemble de ces blocs de questions était présenté une seule fois par position, soit en position 1 (début du test), en positions 2 et 3 et, enfin, en position 4 (fin de test). Comme l'illustre le Tableau 1, si l'on s'intéresse par exemple au domaine de la lecture, chacun des blocs de questions évaluant ce domaine (encadrés en rouge) apparaît une seule fois dans chacune des quatre positions.

Tableau 1  
*Plan d'évaluation du cycle 2012 de PISA*

Carnets de test	Blocs d'items			
	Position 1	Position 2	Position 3	Position 4
1	Math. 5	Scie. 3	Math. 6	Scie. 2
2	Scie. 3	Lect. 3	Math. 7	Lect. 2
3	Lect. 3	Math. 6	Scie. 1	Math. 3
4	Math. 6	Math. 7	Lect. 1	Math. 4
5	Math. 7	Scie. 1	Math. 1	Math. 5
6	Math. 1	Math. 2	Lect. 2	Math. 6
7	Math. 2	Scie. 2	Math. 3	Math. 7
8	Scie. 2	Lect. 2	Math. 4	Scie. 1
9	Lect. 2	Math. 3	Math. 5	Lect. 1
10	Math. 3	Math. 4	Scie. 3	Math. 1
11	Math. 4	Math. 5	Lect. 3	Math. 2
12	Scie. 1	Lect. 1	Math. 2	Scie. 3
13	Lect. 1	Math. 1	Scie. 2	Lect. 3

*Note.* Conception de la rotation des blocs d'items utilisée pour former des livrets de test standard pour PISA 2012 (OCDE, 2014, p. 31).

Les données des carnets de test ont donc été sélectionnées en fonction de la position des blocs de questions et du domaine évalué. À titre d'exemple, pour étudier la persévérance dans le domaine de la lecture, les carnets 3, 9 et 13 ont été retenus pour la position 1 ; les carnets 2, 8 et 12 pour la position 2 ; les carnets 4, 6 et 11 pour la position 3 ; et les carnets 2, 9 et 13 pour la position 4.

Lors du cycle 2015, le plan d'évaluation de PISA a profondément changé :

- a) Le passage à une épreuve informatisée a permis une démultiplication du nombre de livrets, passant de 13 à 36, voire 66 pour les pays ayant opté de participer à l'option internationale de résolution collaborative de problèmes ;
- b) La très grande majorité des élèves ont été évalués dans seulement deux domaines, la première heure étant réservée au premier, et la seconde au deuxième domaine ;
- c) Pour le domaine majeur, à savoir les sciences, un très grand nombre d'items ont été présentés selon un schéma de rotation d'une très grande complexité, de sorte qu'il n'est pas aisé de déterminer si l'item a été présenté en position 1 ou 2 (et respectivement en position 3 ou 4).

Par conséquent, dans le cadre de cet article, la persévérance ne sera pas étudiée dans le domaine des sciences.

Par ailleurs, pour étudier l'effet du mode d'administration, il importe de restreindre cette analyse aux seuls items communs (dits d'ancrage) aux cycles 2012 et 2015. *In fine*, cette étude se base sur 44 questions en lecture et sur 69 questions en mathématiques.

Les réponses à ces items d'ancrage ont été recodées comme suit : la valeur de 1 est attribuée aux réponses correctes, tandis que la valeur de 0 est attribuée aux réponses incorrectes, invalides ou manquantes. Les items non atteints ont également été considérés comme réponses incorrectes, comme ce fut le cas pour les cycles de 2000 à 2012. Enfin, les réponses partiellement correctes pour les items dits à crédits partiels ont été recodées comme réponses incorrectes.

Dans les études précédemment citées, généralement, la persévérance est opérationnalisée comme la différence du taux de réponses correctes observée tout au long du test. Toutefois, cette approche méthodologique ne permet que d'étudier l'évolution de la performance moyenne. Or, puisque la persévérance varie notamment en fonction du niveau de performance de l'élève (Borgonovi et Biecek, 2016), les élèves les plus performants étant plus persévérants, la variabilité de la performance devrait aussi augmenter au fur et à mesure de l'épreuve.

Pour étudier conjointement ces deux effets pour un domaine donné, toutes positions confondues, les réponses des élèves ont, dans un premier temps, été mises à l'échelle selon un modèle logistique de réponse à l'item à un paramètre avec le logiciel ConQuest (Adam et al., 2020). La compétence des élèves dans ce domaine a, dans un second temps, été estimée sous la forme de *weighted likelihood estimations* (WLE; Warm, 1989), séparément pour chaque position, et en ancrant les paramètres de difficultés des items obtenus lors de la première étape. Puisque la plupart des élèves ont été testés pendant une heure pour un domaine donné, on dispose donc, par domaine, pour la plupart des élèves, de deux estimations de la compétence.

Ensuite, les moyennes et les écarts-types ont été calculés, puis les erreurs-types, par pays et par position, en pondérant les données selon la variable reprise dans les bases de données <sup>1</sup> (Ces données sont disponibles à l'Annexe 1 pour le domaine de la lecture en 2012; à l'Annexe 2 pour ce même domaine en 2015; à l'Annexe 3 pour le domaine des mathématiques en 2012; à l'Annexe 4 pour ce domaine en 2015). Les valeurs moyennes observées au sein de l'OCDE ont été obtenues, tous pays confondus, en attribuant à chaque pays le même poids<sup>2</sup>. Les erreurs-types ont été estimées en recourant aux réplifications fournies également dans les bases de données PISA.

## Résultats et interprétation

Le Tableau 2 présente par domaine et par cycle, tous pays confondus, les moyennes et les écarts-types, respectivement en position 1, 2, 3 et 4, pour les domaines de la lecture et des mathématiques. Les erreurs-types figurent entre parenthèses.

Les résultats présentés dans le Tableau 2, plus particulièrement les performances moyennes<sup>3</sup>, tendent à confirmer notre hypothèse 1, en particulier en compréhension de l'écrit. Ainsi, en 2012, la performance moyenne dans ce domaine, au sein des pays de l'OCDE, s'élève à 0,86 en position 1; à 0,64 en position 2; à 0,57 en position 3; et à 0,19 en position 4. En mathématiques, toujours lors du cycle 2012, les moyennes par position s'élèvent respectivement à 0,06; 0,02; -0,12 et -0,32. Les résultats du cycle 2015 se présentent plus en dents de scie, avec néanmoins une tendance à une moindre performance.

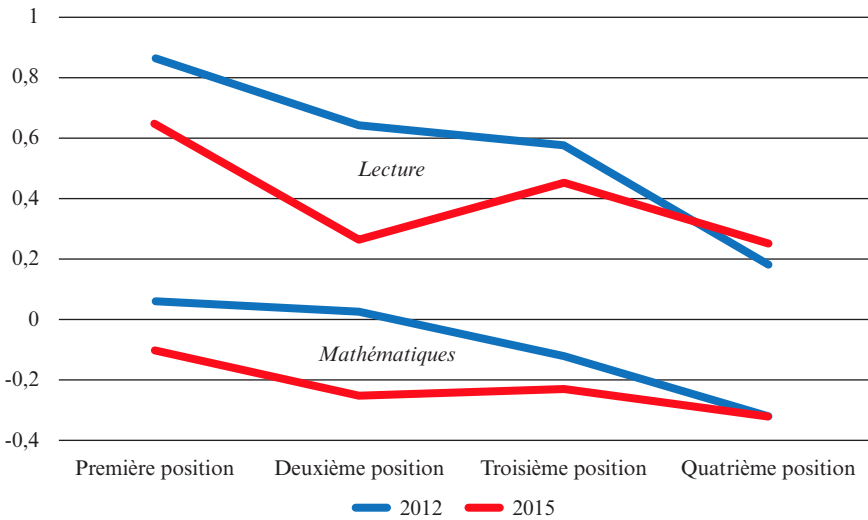
Tableau 2  
*Moyennes et écarts-types de la performance, par position et par domaine*

Domaine	Cycles	Indicateurs statistiques	Position 1		Position 2		Position 3		Position 4	
			Moy.	E. T.	Moy.	E. T.	Moy.	E. T.	Moy.	E. T.
Lecture	2012	$\mu$ (E.T.)	0,86	(0,01)	0,64	(0,01)	0,57	(0,01)	0,19	(0,01)
		$\sigma$ (E.T.)	1,31	(0,01)	1,40	(0,01)	1,42	(0,01)	1,66	(0,01)
	2015	$\mu$ (E.T.)	0,64	(0,01)	0,26	(0,01)	0,45	(0,01)	0,25	(0,01)
		$\sigma$ (E.T.)	1,35	(0,01)	1,45	(0,01)	1,41	(0,01)	1,46	(0,01)
Mathématiques	2012	$\mu$ (E.T.)	0,06	(0,01)	0,02	(0,01)	-0,12	(0,01)	-0,32	(0,01)
		$\sigma$ (E.T.)	1,64	(0,01)	1,66	(0,01)	1,70	(0,01)	1,77	(0,01)
	2015	$\mu$ (E.T.)	-0,11	(0,01)	-0,25	(0,01)	-0,23	(0,01)	-0,32	(0,01)
		$\sigma$ (E.T.)	1,61	(0,01)	1,63	(0,02)	1,64	(0,01)	1,66	(0,01)

Ces variations sont représentées graphiquement dans la Figure 1. En compréhension de l'écrit pour le cycle 2015, on observe bien une diminution de la performance entre les positions 1 et 2 ainsi qu'entre les positions 3 et 4. Toutefois, la performance moyenne en position 3 est plus élevée qu'en position 2. Par contre, l'évolution de la performance en mathématiques n'est pas caractérisée par ce rebond inattendu.

Le changement de design lors du cycle 2015 pourrait expliquer ce rebond inattendu en compréhension de l'écrit. En effet, comme susmentionné, lors de ce cycle, la très grande majorité des élèves ont été évalués dans un domaine durant la première heure et dans un second domaine durant la seconde heure, alors qu'en 2012 près de la moitié des élèves ont été évalués dans trois domaines. Dès lors, travailler sur un même domaine pendant une heure aurait pu entraîner, notamment par lassitude, une diminution plus importante de la persévérance, très visible en lecture, perceptible en mathématiques. La pause entre les deux heures aurait quant à elle permis, d'une certaine manière, de restimuler (*reboost*) la persévérance des élèves, particulièrement pour les élèves ayant été évalués pendant la première heure en mathématiques ou en sciences et lors de la seconde heure en compréhension de l'écrit.

Figure 1  
*Performances moyennes, par position, par domaine et par cycle*



Afin de mieux comprendre ce résultat plutôt surprenant en lecture en 2015, les performances moyennes par position et par genre ont été calculées pour ce domaine (voir Tableau 3). En effet, comme le suggère la littérature scientifique, les filles ont des croyances plus positives que les garçons pour les activités de lecture (Eccles et al., 1993), sont en moyenne plus engagées que les garçons dans la lecture (OCDE, 2002) et démontrent toujours en lecture une plus grande persévérance que les garçons (Borgonovi et Biecek, 2016). Ces observations scientifiques laissent présager que ce regain de persévérance observé après la pause et après le changement de domaine en 2015 serait particulièrement prononcé chez les filles. Ainsi, après avoir passé une heure à répondre à des questions de sciences ou de mathématiques, elles seraient, d'une certaine manière, plus motivées à répondre à des questions de compréhension à l'écrit.

Les données reprises dans le Tableau 3 semblent le confirmer partiellement. En effet, aucune augmentation de persévérance n'est perceptible entre les positions 2 et 3 lors du cycle 2012, quel que soit le genre des élèves. Au contraire, comme susmentionné, la tendance est à la baisse, tant pour les filles que pour les garçons. Par contre, en 2015, la différence

Tableau 3  
*Moyennes et écarts-types de la performance en lecture lors des cycles 2012 et 2015, par position et par genre*

Genre	Cycles	Indicateurs statistiques	Position 1		Position 2		Position 3		Position 4	
			Moy.	E. T.	Moy.	E. T.	Moy.	E. T.	Moy.	E. T.
Filles	2012	$\mu$ (E.T.)	1,02	(0,01)	0,85	(0,01)	0,78	(0,01)	0,44	(0,01)
		$\sigma$ (E.T.)	1,26	(0,01)	1,30	(0,01)	1,32	(0,01)	1,56	(0,01)
	2015	$\mu$ (E.T.)	0,77	(0,01)	0,40	(0,02)	0,63	(0,01)	0,42	(0,01)
		$\sigma$ (E.T.)	1,31	(0,01)	1,43	(0,01)	1,34	(0,01)	1,41	(0,01)
Garçons	2012	$\mu$ (E.T.)	0,70	(0,01)	0,44	(0,01)	0,36	(0,01)	-0,05	(0,01)
		$\sigma$ (E.T.)	1,34	(0,01)	1,46	(0,01)	1,48	(0,01)	1,71	(0,01)
	2015	$\mu$ (E.T.)	0,51	(0,01)	0,13	(0,02)	0,27	(0,01)	0,08	(0,01)
		$\sigma$ (E.T.)	1,37	(0,01)	1,46	(0,01)	1,45	(0,01)	1,49	(0,01)

de performance entre les positions 2 et 3 est plus accentuée pour les filles (0,40 - 0,63 = -0,23) que pour les garçons (0,13 - 0,27 = -0,14), ce qui pourrait être expliqué par le changement de plan d'évaluation lors de ce cycle.

Concernant notre hypothèse 2, les résultats présentés dans le Tableau 2 soulignent que les erreurs-types tendent à augmenter à travers les quatre positions, et ce, quel que soit le domaine. Cette augmentation de la variabilité est plus élevée en compréhension de l'écrit qu'en mathématiques. Par ailleurs, elle est nettement plus importante pour la version papier-crayon de 2012 que pour la version informatisée de 2015. Ces résultats confirment donc l'hypothèse 2.

La possibilité offerte par l'approche adoptée dans cet article permet également de traduire les évolutions de la persévérance selon la position des items sous la forme d'ampleurs de l'effet. Dans le cadre de cette étude, elle est égale à :  $\frac{\mu_{p1} - \mu_{p4}}{\sqrt{\frac{\sigma_{p1}^2 + \sigma_{p4}^2}{2}}}$ .

Le Tableau 4 présente les ampleurs de l'effet, par domaine et par cycle.



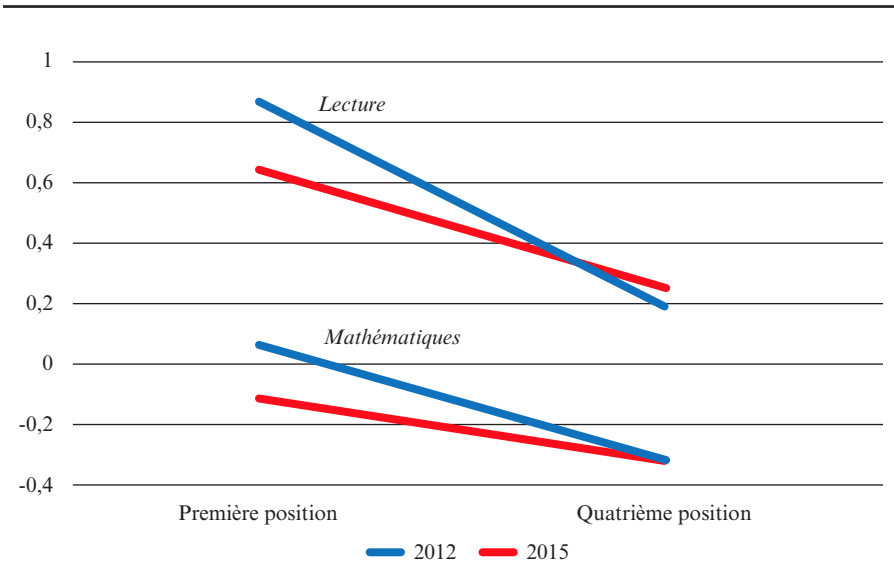
Tableau 4  
*Ampleurs de l'effet, par domaine et par cycle*

Domaine	Cycles	Ampleurs de l'effet
Lecture	2012	0,44
	2015	0,27
Mathématiques	2012	0,22
	2015	0,13

Les résultats présentés dans le Tableau 4 permettent à nouveau de corroborer notre hypothèse 1. En effet, les ampleurs de l'effet suggèrent que la performance moyenne des élèves a effectivement diminué entre les positions 1 et 4. En outre, les tailles de l'effet sont plus importantes dans le domaine de la lecture qu'en mathématiques, ce qui semble confirmer notre hypothèse 3. En effet, les ampleurs de l'effet en lecture sont respectivement de 0,44 en 2012 et de 0,27 en 2015, alors qu'elles ne sont que de 0,22 en 2012 et de 0,13 en 2015 en mathématiques.

Ces résultats, représentés graphiquement à la Figure 2, suggèrent que la diminution de persévérance est moins prononcée en mathématiques qu'en lecture. Comme l'expliquent Braun et al. (2011), la lecture nécessite une

Figure 2  
*Performances moyennes en position 1 et 4, par domaine et par cycle*



attention plus importante que les autres domaines. De telles variations laissent donc présager que le domaine pourrait «influencer» la persévérance des élèves. Notons toutefois que les analyses conduites dans le cadre de cette étude ne permettent pas de confirmer la causalité du domaine évalué.

Les résultats présentés dans le Tableau 4 semblent également confirmer notre hypothèse 4. En effet, les ampleurs de l'effet sont plus faibles au cycle 2015 qu'en 2012. Ce constat pourrait donc confirmer une probable influence du mode d'administration et coïnciderait avec les observations sur le rapport de l'OCDE de 2010 faites par Dierendonck et al. (2013), qui suggèrent que les élèves déclarent être plus motivés à répondre au test informatisé qu'au test papier-crayon. De même, Khoshima et Hashemi (2017) suggèrent que les élèves ont mieux performé lors des tests électroniques, mais que cet effet pourrait varier en fonction du domaine évalué (Kingston, 2008).

Cela dit, les différences de performance représentées à la Figure 2 entre les positions 1 et 4 selon le cycle d'évaluation suggèrent qu'en moyenne les élèves ont mieux répondu aux questions posées en position 1 lors du cycle 2012, comparativement à ces mêmes questions posées en position 1 lors du cycle 2015. Par contre, quand ces mêmes items étaient posés en position 4, les différences de performance s'estompent, voire s'inversent : les élèves ont légèrement mieux répondu en 2015 qu'en 2012 en lecture et de manière équivalente en mathématiques. Plus précisément, en lecture, la différence de moyennes entre les deux cycles s'élève à 0,22 (0,86 - 0,64) en position 1 et à -0,06 (0,19 - 0,25) en position 4. Ce constat est également observé en mathématiques, où la différence de moyennes entre les deux cycles s'élève à 0,17 (0,06 - (-0,11)) en position 1 et à 0,00 (0,32 - (-0,32)) en position 4. En outre, les pentes des droites représentant le cycle 2015 sont moins inclinées que celles représentant le cycle 2012, ce qui laisse supposer que les élèves se sont montrés plus persévérants lors du test informatisé, ou inversement (que les élèves ont moins persévéré lors du test papier-crayon).

Comment peut-on expliquer ce phénomène ? Peu ou pas habitués à répondre à un test informatisé, les élèves éprouveraient une charge cognitive plus importante en début de test pour le format électronique, ce qui engendrerait cette performance moindre, comparativement au test papier-crayon. Une fois coutumiers de ce nouvel environnement, ils se laisseraient «embarquer» par le caractère novateur de cette évaluation.

Ces différents résultats interpellent donc la comparabilité des résultats entre les deux cycles et suggèrent qu'un test informatisé ne devrait pas être considéré comme la simple numérisation d'un test papier-crayon. En effet, comme le rappellent Ebrahimi et al. (2019), «un test est fiable lorsqu'il mesure régulièrement ce qu'il est censé mesurer en produisant des scores stables et constants à deux occasions de test» (p. 131, trad. libre).

## Discussion

Cette recherche avait pour objet d'étudier la persévérance des élèves lors des évaluations à faibles enjeux, en particulier l'effet du mode d'administration et du domaine évalué sur celle-ci. L'enquête PISA, en implantant un plan d'évaluation incomplet balancé, offre un matériel de prédilection pour analyser cette persévérance. La présente étude a exploité les données des cycles 2012 et 2015 en opérationnalisant la persévérance comme l'évolution de la performance tout au long de l'épreuve.

Les quatre hypothèses susmentionnées tendent à être confirmées. Tout d'abord, si l'on observe bien des performances plus élevées en début de test qu'en fin de test, quels que soient le cycle ou le domaine évalué, le différentiel observé est plus élevé lors du cycle 2012, ce qui laisse sous-entendre, toutes choses égales par ailleurs, que les élèves seraient plus persévérants avec un test informatisé.

Si les élèves semblent davantage maintenir leur motivation à répondre à une épreuve numérisée, ils semblent aussi avoir besoin d'un petit temps d'adaptation aux plateformes informatiques. En effet, avec ce nouvel environnement, les performances observées en début de test sont inférieures aux performances observées avec une épreuve papier-crayon. L'informatisation du test pourrait en quelque sorte engendrer une charge cognitive supplémentaire, particulièrement en début de test, qui se concrétiserait par une moindre performance. Par contre, ce déclin en début d'épreuve est compensé par une plus grande persévérance, de sorte qu'en fin d'épreuve les performances à la version informatisée se rapprochent de celles observées au mode papier-crayon. Autrement dit, avec un test papier-crayon, les répondants débutent mieux, mais «s'essoufflent» plus fortement. Comme dirait Jean de La Fontaine, *rien ne sert de courir, il faut partir à point*.

Enfin, quel que soit le cycle et/ou le mode d'administration, la persévérance des élèves fluctue en fonction du domaine évalué. Globalement, les élèves démontrent une plus grande persévérance en mathématiques et une plus faible en compréhension de l'écrit. Il serait utile d'étudier au sein d'un domaine si la persévérance fluctue avec la longueur de la mise en contexte (soit le texte que l'élève doit lire avant de prendre connaissance de la question/des questions s'y rapportant). Une persévérance supérieure pour les unités avec de plus petites amorces pourrait notamment expliquer les différences observées entre les mathématiques et la compréhension de l'écrit.

### Perspectives et limites

L'originalité majeure de cette recherche est qu'elle associe mesure de la persévérance et analyse de l'effet du mode d'administration, ce qui, à notre connaissance, n'avait pas encore été étudié ; du moins, sur des échantillons représentatifs de grande taille dans 33 pays. Nul doute que les résultats rapportés alimenteront les débats liés à la motivation des répondants lors des évaluations à faibles enjeux ou à l'effet du mode d'administration. Malgré cet apport indéniable, plusieurs limites sont à souligner.

La première résulte directement de la définition adoptée de la *persévérance* dans le cadre de cette étude. Toutes choses étant égales par ailleurs, et à l'instar des critiques susmentionnées à l'égard de l'indice d'effort relatif de Butler et Adams (2007), une probabilité constante de répondre correctement tout au long de l'épreuve peut aussi traduire un manque uniforme de motivation. En effet, des élèves pourraient être considérés comme persévérants sans nécessairement être motivés à répondre correctement, pourvu qu'ils le fassent de manière équivalente de la position 1 à la position 4.

De plus, dans le cadre de cette étude, les omissions ont été considérées comme des réponses incorrectes. L'étude de la persévérance pourrait tout autant se concentrer exclusivement sur les taux d'items non atteints. Néanmoins, deux raisons justifient la méthodologie retenue. D'une part, selon nous, les omissions ne constituent qu'une partie du problème. En effet, un élève peut être moins performant, tout en continuant à répondre. Par conséquent, on pourrait observer une diminution de la performance, sans pour autant qu'il y ait une augmentation des omissions. Limiter l'analyse à ces seuls items non atteints pourrait donc conduire éventuellement à une sous-estimation du déclin de la performance observée. D'autre part,

une omission à un item peut traduire un manque de temps, de connaissance ou de motivation du répondant. Il est en effet possible que les élèves aient eu moins de temps pour répondre aux questions posées en fin de test, comparativement à celles posées en début de test. Un taux d'omissions plus important en position 4 pourrait donc être dû à un manque de temps, et non de persévérance.

Par ailleurs, les différences de performance en fonction de la position des questions peuvent être imputées à un manque de persévérance, mais également à la fatigue, sans que nous puissions différencier ces différentes causes. Par exemple, un élève fatigué serait très probablement plus distrait et commettrait davantage d'erreurs techniques, même en restant motivé (et persévérant). Il est donc légitime de se poser la question de savoir si, à persévérance égale, mais de fatigue différente, les élèves auraient répondu de manière identique ou non. On ne peut également pas déterminer si le faible enjeu du test a une influence sur cette fatigue ou ce manque de persévérance. En effet, aucune donnée sur les profils motivationnels des élèves n'a été utilisée en parallèle dans la présente étude, notamment parce qu'aucune donnée auto-rapportée d'effort n'était disponible pour le cycle 2015. Sans l'apport d'autres mesures telles que le thermomètre à l'effort, la confusion subsistera. Une autre option serait d'élaborer une étude expérimentale, voire quasi-expérimentale qui permettrait d'analyser la persévérance (ou la fatigue) en fonction des enjeux du test et de sa longueur.

Enfin, le mode d'administration ne constitue pas la seule et unique différence entre les plans d'évaluation de 2012 et de 2015. Ainsi, en 2015, la très grande majorité des élèves n'ont été évalués que dans deux domaines, le premier durant la première heure et le second durant la seconde heure, de sorte que les élèves n'étaient confrontés qu'à un seul changement de domaine. En 2012, 7/13 des élèves ont connu deux changements et 6/13 trois changements. Or, le changement de domaines au sein d'un livret peut entraîner une augmentation de la charge cognitive des répondants. Par conséquent, les différences de comportement observées entre ces deux cycles ne peuvent pas être imputées uniquement au changement de mode d'administration.

Réception : 8 janvier 2021

Version finale : 8 novembre 2021

Acceptation : 3 janvier 2022

## NOTES

1. Sur base de la variable W\_FSTUWT.
2. Utilisation des *SENAT weight*.
3. Les scores TRI peuvent varier approximativement entre -4 et +4. Une moyenne légèrement négative signifie que le taux moyen de réussite est inférieur à 50%.

## RÉFÉRENCES

- Adams, R. J., Wu, M. L., Cloney, D., & Wilson, M. R. (2020). *ACER ConQuest : Generalised item response modelling software* (Version 5) [Computer software]. Camberwell: Australian Council for Educational Research.
- Asseburg, R., & Frey, A. (2013). Too hard, too easy, or just right? The relationship between effort or boredom and ability-difficulty fit. *Psychological Test and Assessment Modeling*, 55(1), 92-104. [https://www.researchgate.net/publication/267907815\\_Too\\_hard\\_too\\_easy\\_or\\_just\\_right\\_The\\_relationship\\_between\\_effort\\_or\\_boredom\\_and\\_ability-difficulty\\_fit](https://www.researchgate.net/publication/267907815_Too_hard_too_easy_or_just_right_The_relationship_between_effort_or_boredom_and_ability-difficulty_fit)
- Borgonovi, F., & Biecek, P. (2016). An international comparison of students' ability to endure fatigue and maintain motivation during a low-stakes test. *Learning and Individual Differences*, 49, 128-137. <https://doi.org/10.1016/j.lindif.2016.06.001>
- Braun, H., Kirsch, I., & Yamamoto, K. (2011). An experimental study of the effects of monetary incentives on performance on the 12th-grade NAEP reading assessment. *Teachers College Record*, 113(11), 2309-2344. <https://doi.org/10.1177/2F016146811111301101>
- Butler, J., & Adams, R. J. (2007). The impact of differential investment of student effort on the outcomes of international studies. *Journal of Applied Measurement*, 8(3), 279-304. [https://www.researchgate.net/publication/6035437\\_The\\_impact\\_of\\_differential\\_investment\\_of\\_student\\_effort\\_on\\_the\\_outcomes\\_of\\_international\\_studies](https://www.researchgate.net/publication/6035437_The_impact_of_differential_investment_of_student_effort_on_the_outcomes_of_international_studies)
- Chen, P. P. (2002). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences*, 14, 77-90. <https://doi.org/10.1016/j.lindif.2003.08.003>
- Debeer, D., Buchholz, J., Hartig, J., & Janssen, R. (2014). Student, school, and country differences in sustained test-taking effort in the 2009 PISA reading assessment. *Journal of Educational and Behavioral Statistics*, 39(6), 502-523. <https://doi.org/10.3102%2F1076998614558485>
- Dierendonck, C., Burton, R. & Baye, A. (2013). *Effort consenti par les élèves et performance aux évaluations externes à faibles enjeux : illustration au départ des données PISA* [communication orale]. Congrès de l'Actualité de la recherche en éducation et formation (AREF), Montpellier, France. [https://orbi.uliege.be/bitstream/2268/196025/1/AREF\\_communication%202\\_Baye.pdf](https://orbi.uliege.be/bitstream/2268/196025/1/AREF_communication%202_Baye.pdf)
- Ebrahimi, M. R., Toroujeni, S. M. H., & Shahbazi, V. (2019). Score equivalence, gender difference, and testing mode preference in a comparative study between computer-based testing and paper-based testing. *International Journal of Emerging Technologies in Learning*, 14(7), 128-143. <https://doi.org/10.3991/ijet.v14i07.10175>

- Eccles, J., Wigfield, A., Harold, R., & Blumenfeld, P. (1993). Age and gender differences in children's self and task perceptions during elementary school. *Child Development*, 64, 830-847. <https://doi.org/10.1111/j.1467-8624.1993.tb02946.x>
- Eklöf, H. (2007). Test-taking motivation and mathematics performance in TIMSS 2003. *International Journal of Testing*, 7(3), 311-326. <https://doi.org/10.1080/15305050701438074>
- Eklöf, H. (2015). Swedish students' reported motivation and effort in PISA, over time and in comparison with other countries. In Swedish National Agency for Education, *To respond or not to respond: The motivation of Swedish students in taking the PISA test* (p. 11-60). Swedish National Agency for Education.
- Eklöf, H., & Nyroos, M. (2013). Pupil perceptions of national tests in science: Perceived importance, invested effort, and test anxiety. *European Journal of Psychology of Education*, 28(2), 497-510. <https://doi.org/10.1007/s10212-012-0125-6>
- Eklöf, H., Pavešič, B. J., & Grønmo, L. S. (2014). Cross-national comparison of reported effort and mathematics performance in TIMSS advanced. *Applied Measurement in Education*, 27(1), 31-45. <https://doi.org/10.1080/08957347.2013.853070>
- Fumel, S. & Keskpaik, S. (2017). La motivation des élèves à répondre à un test standardisé : résultats d'une étude dans le cadre de CEDRE compétences langagières et littératie. *Éducation & formations*, 93, 105-119. <https://halshs.archives-ouvertes.fr/halshs-01577258/document>
- Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales: The reference-group problem. *Journal of Personality and Social Psychology*, 82, 903-918. <https://doi.apa.org/doi/10.1037/0022-3514.82.6.903>
- Hohensinn, C., Kubinger, K. D., Reif, M., Schleicher, E. & Khorramdel, L. (2011). Analysing item position effects due to test booklet design within large-scale assessment. *Educational Research and Evaluation*, 17(6), 497-509. <https://doi.org/10.1080/13803611.2011.632668>
- Hopfenbeck, T. N., & Kjærnsli, M. (2016). Students' test motivation in PISA: The case of Norway. *The Curriculum Journal*, 27(3), 406-422. <https://doi.org/10.1080/09585176.2016.1156004>
- Keskpaik, S. & Rocher, T. (2015). La motivation des élèves français face à des évaluations à faibles enjeux. Comment la mesurer? Son impact sur les réponses. *Éducation & formations*, 86-87, 119-139. [https://www.researchgate.net/publication/316990893\\_La\\_motivation\\_des\\_eleves\\_francais\\_face\\_a\\_des\\_evaluations\\_a\\_faibles\\_enjeux\\_Comment\\_la\\_mesurer\\_Son\\_impact\\_sur\\_les\\_reponses](https://www.researchgate.net/publication/316990893_La_motivation_des_eleves_francais_face_a_des_evaluations_a_faibles_enjeux_Comment_la_mesurer_Son_impact_sur_les_reponses)
- Khoshsima, H., & Hashemi, S. M. T. (2017). Transitioning to an alternative assessment: Computer-based testing and key factors related to testing mode. *European Journal of English Language Teaching*, 2(1), 54-74. <https://doi.org/10.5281/zenodo.268576>
- Kingston, N. M. (2008) Comparability of computer- and paper-administered multiple-choice tests for k-12 populations: A synthesis. *Applied Measurement in Education*, 22(1), 22-37. <https://doi.org/10.1080/08957340802558326>
- List, J. A., Gneezy, U., Livingston, J. A., Qin, X., Sadoff, S., & Xu, Y. (2019). *Measuring success in education: The role of effort on the test itself*. [Working paper no. 66]. Becker Friedman Institute. <https://ssrn.com/abstract=3386365>

- Nagy, G., Nagengast, B., Frey, A., Becker, M., & Rose, N. (2019). A multilevel study of position effects in PISA achievement tests: Student- and school-level predictors in the German tracked school system. *Assessment in Education: Principles, Policy & Practice*, 26(4), 422-443. <https://doi.org/10.1080/0969594X.2018.1449100>
- O'Neil, H. F., Abedi, J., Miyoshi, J., & Mastergeorge, A. (2005). Monetary incentives for low-stakes tests. *Educational Assessment*, 10(3), 185-208. [https://doi.org/10.1207/s15326977ea1003\\_3](https://doi.org/10.1207/s15326977ea1003_3)
- O'Neil, H. F., Sugrue, B., & Baker, E. L. (1996). Effects of motivational interventions on the national assessment of educational progress mathematics performance. *Educational Assessment*, 2(2), 135-157. [https://doi.org/10.1207/s15326977ea0302\\_2](https://doi.org/10.1207/s15326977ea0302_2)
- Organisation de coopération et de développement économiques (OCDE). (2002). *La lecture, moteur de changement : performances et engagement d'un pays à l'autre – Résultats du cycle d'enquêtes de PISA 2000*. Éditions OCDE. <https://doi.org/10.1787/9789264099296-fr>
- Organisation de coopération et de développement économiques (OCDE). (2016). *Cadre d'évaluation et d'analyse de l'enquête PISA 2015 : compétences en sciences, en compréhension de l'écrit, en mathématiques, en matières financières et en résolution collaborative de problèmes (éd. révisée)*. Éditions OCDE. <https://www.oecd-ilibrary.org/docserver/9789264297203-fr.pdf?expires=1643212640&id=id&accname=guest&checksum=1173D79592CBF7B633EAB4B68902A086>
- Organisation for Economic Co-operation and Development (OECD). (2014). *PISA 2012 : Technical report*. OECD Editions. <https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf>
- Organisation for Economic Co-operation and Development (OECD). (2017). *PISA 2015 : Technical report*. OECD Editions. <https://www.oecd.org/pisa/data/2015-technical-report>
- Penk, C., & Schipolowski, S. (2015). Is it all about value? Bringing back the expectancy component to the assessment of test-taking motivation. *Learning and Individual Differences*, 42, 27-35. <http://dx.doi.org/10.1016/j.lindif.2015.08.002>
- Steedle, J. T., & Grochowalski, J. (2017). The effect of stakes on accountability test scores and pass rates. *Educational Assessment*, 22(2), 111-123. <https://doi.org/10.1080/10627197.2017.1309276>
- Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2007). A meta-analysis of testing mode effects in grade K-12 mathematics tests. *Educational and Psychological Measurement*, 67(2), 219-238. <https://doi.org/10.1177%2F0013164406288166>
- Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2008). Comparability of computer-based and paper-and-pencil testing in K-12 reading assessments: A meta-analysis of testing mode effects. *Educational and Psychological Measurement*, 68(1), 5-24. <https://doi.org/10.1177%2F0013164407305592>
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3), 427-450. <https://doi.org/10.1007/BF02294627>
- Wigfield, A., & Eccles, J. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25, 68-81. <https://doi.org/10.1006/ceps.1999>
- Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, 10(1), 1-17. [https://doi.org/10.1207/s15326977ea1001\\_1](https://doi.org/10.1207/s15326977ea1001_1)



## Annexe 1

*Moyennes et écarts-types de la performance des élèves, par pays et par position, en lecture au cycle 2012*

Pays	Position 1				Position 2				Position 3				Position 4			
	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.
AUS	0,94	(0,03)	1,32	(0,02)	0,75	(0,03)	1,38	(0,02)	0,71	(0,03)	1,43	(0,02)	0,42	(0,04)	1,58	(0,03)
AUT	0,59	(0,04)	1,27	(0,03)	0,50	(0,05)	1,37	(0,03)	0,50	(0,05)	1,32	(0,04)	0,19	(0,05)	1,53	(0,05)
BEL	0,99	(0,03)	1,29	(0,03)	0,78	(0,04)	1,38	(0,03)	0,77	(0,03)	1,45	(0,03)	0,42	(0,04)	1,70	(0,03)
CAN	1,08	(0,03)	1,24	(0,02)	0,92	(0,03)	1,36	(0,02)	0,85	(0,03)	1,41	(0,03)	0,44	(0,04)	1,67	(0,04)
CHE	0,90	(0,03)	1,22	(0,03)	0,69	(0,05)	1,35	(0,03)	0,70	(0,03)	1,37	(0,03)	0,35	(0,04)	1,58	(0,04)
CZE	0,85	(0,04)	1,23	(0,03)	0,59	(0,05)	1,34	(0,03)	0,54	(0,04)	1,29	(0,04)	0,24	(0,05)	1,48	(0,05)
DEU	0,91	(0,04)	1,22	(0,03)	0,84	(0,05)	1,34	(0,03)	0,65	(0,05)	1,40	(0,04)	0,40	(0,06)	1,47	(0,04)
DNK	0,80	(0,05)	1,22	(0,03)	0,54	(0,05)	1,26	(0,04)	0,59	(0,04)	1,26	(0,05)	0,29	(0,05)	1,44	(0,05)
ESP	0,74	(0,03)	1,32	(0,02)	0,44	(0,04)	1,34	(0,03)	0,44	(0,03)	1,34	(0,02)	-0,08	(0,04)	1,60	(0,03)
EST	1,02	(0,04)	1,18	(0,03)	0,90	(0,04)	1,20	(0,03)	0,73	(0,05)	1,26	(0,03)	0,53	(0,05)	1,37	(0,04)
FIN	1,06	(0,04)	1,28	(0,03)	0,95	(0,04)	1,33	(0,03)	0,86	(0,04)	1,40	(0,03)	0,70	(0,04)	1,56	(0,04)
FRA	0,97	(0,04)	1,37	(0,03)	0,72	(0,05)	1,56	(0,05)	0,64	(0,05)	1,66	(0,05)	0,16	(0,06)	1,92	(0,06)
GBR	0,80	(0,04)	1,30	(0,03)	0,61	(0,05)	1,39	(0,04)	0,53	(0,05)	1,46	(0,03)	0,18	(0,07)	1,69	(0,05)
GRC	0,77	(0,05)	1,36	(0,05)	0,41	(0,05)	1,46	(0,04)	0,21	(0,05)	1,42	(0,04)	-0,59	(0,06)	1,75	(0,04)
HUN	0,69	(0,04)	1,25	(0,03)	0,51	(0,06)	1,44	(0,04)	0,47	(0,06)	1,44	(0,04)	0,01	(0,06)	1,60	(0,04)
IRL	1,04	(0,03)	1,21	(0,03)	0,89	(0,05)	1,30	(0,04)	0,86	(0,05)	1,28	(0,04)	0,63	(0,05)	1,49	(0,04)
ISL	0,68	(0,05)	1,34	(0,05)	0,37	(0,06)	1,48	(0,05)	0,31	(0,05)	1,45	(0,04)	-0,17	(0,06)	1,62	(0,04)
ISR	0,77	(0,06)	1,44	(0,04)	0,56	(0,07)	1,62	(0,04)	0,33	(0,08)	1,65	(0,05)	-0,30	(0,09)	1,92	(0,05)

Pays	Position 1			Position 2			Position 3			Position 4						
	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.
<b>ITA</b>	0,76	(0,02)	1,29	(0,02)	0,53	(0,03)	1,45	(0,02)	0,40	(0,03)	1,45	(0,02)	-0,03	(0,04)	1,71	(0,02)
<b>JPN</b>	1,41	(0,05)	1,29	(0,03)	1,01	(0,05)	1,41	(0,04)	1,04	(0,05)	1,45	(0,04)	0,83	(0,06)	1,67	(0,04)
<b>KOR</b>	1,25	(0,05)	1,18	(0,03)	0,99	(0,05)	1,24	(0,03)	0,97	(0,07)	1,38	(0,04)	0,83	(0,06)	1,44	(0,04)
<b>LUX</b>	0,76	(0,04)	1,43	(0,03)	0,50	(0,04)	1,54	(0,04)	0,41	(0,04)	1,52	(0,03)	-0,07	(0,05)	1,79	(0,04)
<b>LVA</b>	0,63	(0,04)	1,19	(0,03)	0,49	(0,05)	1,27	(0,04)	0,47	(0,05)	1,26	(0,04)	0,01	(0,06)	1,44	(0,04)
<b>NLD</b>	1,03	(0,05)	1,25	(0,04)	0,83	(0,05)	1,29	(0,04)	0,69	(0,06)	1,31	(0,06)	0,49	(0,06)	1,46	(0,05)
<b>NOR</b>	0,94	(0,05)	1,37	(0,04)	0,69	(0,06)	1,42	(0,06)	0,66	(0,05)	1,41	(0,04)	0,11	(0,08)	1,82	(0,05)
<b>NZL</b>	1,05	(0,05)	1,41	(0,04)	0,77	(0,05)	1,48	(0,04)	0,64	(0,05)	1,48	(0,04)	0,38	(0,06)	1,78	(0,05)
<b>POL</b>	1,04	(0,05)	1,27	(0,03)	0,91	(0,05)	1,28	(0,03)	0,78	(0,04)	1,28	(0,04)	0,41	(0,06)	1,46	(0,04)
<b>PRT</b>	0,81	(0,05)	1,33	(0,04)	0,51	(0,06)	1,36	(0,04)	0,37	(0,06)	1,40	(0,04)	-0,05	(0,06)	1,64	(0,04)
<b>SVK</b>	0,39	(0,05)	1,40	(0,04)	0,24	(0,06)	1,53	(0,05)	0,20	(0,06)	1,46	(0,05)	-0,31	(0,06)	1,65	(0,04)
<b>SVN</b>	0,61	(0,04)	1,20	(0,02)	0,42	(0,04)	1,35	(0,03)	0,29	(0,05)	1,32	(0,03)	0,07	(0,04)	1,51	(0,04)
<b>SWE</b>	0,72	(0,05)	1,43	(0,04)	0,48	(0,06)	1,45	(0,04)	0,34	(0,05)	1,52	(0,04)	-0,21	(0,07)	1,88	(0,04)
<b>TUR</b>	0,47	(0,06)	1,21	(0,03)	0,34	(0,05)	1,25	(0,03)	0,31	(0,07)	1,30	(0,04)	-0,20	(0,06)	1,48	(0,04)
<b>USA</b>	0,78	(0,05)	1,24	(0,04)	0,52	(0,06)	1,38	(0,03)	0,53	(0,05)	1,31	(0,04)	0,27	(0,07)	1,57	(0,04)

Note. Les données présentées entre parenthèses sont les erreurs-types.

## Annexe 2

*Moyennes et écarts-types de la performance des élèves, par pays et par position, en lecture au cycle 2015*

Pays	Position 1				Position 2				Position 3				Position 4			
	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.
AUS	0,66	(0,04)	1,36	(0,03)	0,32	(0,05)	1,52	(0,03)	0,44	(0,04)	1,41	(0,03)	0,19	(0,04)	1,51	(0,03)
AUT	0,54	(0,05)	1,27	(0,04)	0,20	(0,05)	1,38	(0,04)	0,30	(0,06)	1,46	(0,05)	0,13	(0,05)	1,50	(0,04)
BEL	0,83	(0,04)	1,32	(0,03)	0,44	(0,05)	1,49	(0,04)	0,54	(0,05)	1,42	(0,03)	0,35	(0,05)	1,52	(0,04)
CAN	0,89	(0,04)	1,30	(0,03)	0,54	(0,05)	1,40	(0,03)	0,69	(0,04)	1,33	(0,03)	0,50	(0,05)	1,38	(0,03)
CHE	0,62	(0,06)	1,34	(0,04)	0,30	(0,07)	1,47	(0,04)	0,39	(0,06)	1,44	(0,03)	0,13	(0,06)	1,51	(0,04)
CZE	0,60	(0,05)	1,30	(0,05)	0,13	(0,05)	1,34	(0,04)	0,38	(0,05)	1,33	(0,04)	0,14	(0,06)	1,38	(0,04)
DEU	0,74	(0,05)	1,34	(0,04)	0,41	(0,06)	1,48	(0,05)	0,68	(0,05)	1,36	(0,04)	0,53	(0,06)	1,36	(0,04)
DNK	0,65	(0,06)	1,34	(0,04)	0,36	(0,08)	1,41	(0,05)	0,49	(0,06)	1,36	(0,04)	0,32	(0,05)	1,44	(0,05)
ESP	0,63	(0,05)	1,31	(0,03)	0,16	(0,06)	1,43	(0,05)	0,35	(0,06)	1,29	(0,03)	0,11	(0,06)	1,41	(0,04)
EST	0,87	(0,05)	1,36	(0,04)	0,53	(0,06)	1,43	(0,05)	0,68	(0,06)	1,31	(0,03)	0,55	(0,06)	1,34	(0,04)
FIN	0,91	(0,06)	1,36	(0,04)	0,68	(0,06)	1,35	(0,04)	0,89	(0,06)	1,36	(0,04)	0,70	(0,06)	1,37	(0,04)
FRA	0,74	(0,06)	1,48	(0,05)	0,27	(0,06)	1,53	(0,05)	0,42	(0,06)	1,59	(0,05)	0,27	(0,06)	1,63	(0,05)
GBR	0,66	(0,05)	1,34	(0,04)	0,40	(0,05)	1,43	(0,04)	0,54	(0,06)	1,33	(0,03)	0,34	(0,05)	1,41	(0,04)
GRC	0,41	(0,07)	1,27	(0,04)	-0,19	(0,07)	1,37	(0,06)	0,14	(0,07)	1,47	(0,06)	-0,06	(0,09)	1,50	(0,08)
HUN	0,41	(0,06)	1,34	(0,04)	-0,08	(0,06)	1,49	(0,05)	0,17	(0,05)	1,33	(0,04)	-0,02	(0,06)	1,38	(0,04)
IRL	0,68	(0,05)	1,24	(0,03)	0,54	(0,05)	1,35	(0,05)	0,74	(0,05)	1,27	(0,03)	0,60	(0,05)	1,34	(0,04)
ISL	0,47	(0,06)	1,35	(0,05)	0,08	(0,07)	1,45	(0,06)	0,25	(0,08)	1,40	(0,06)	0,29	(0,09)	1,47	(0,05)
ISR	0,30	(0,07)	1,58	(0,04)	-0,14	(0,05)	1,59	(0,04)	0,19	(0,06)	1,63	(0,05)	-0,01	(0,07)	1,60	(0,04)

Pays	Position 1			Position 2			Position 3			Position 4						
	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.
ITA	0,53	(0,05)	1,28	(0,04)	0,05	(0,06)	1,41	(0,04)	0,29	(0,05)	1,37	(0,05)	0,08	(0,07)	1,52	(0,06)
JPN	0,99	(0,06)	1,34	(0,04)	0,42	(0,06)	1,51	(0,04)	0,94	(0,06)	1,43	(0,05)	0,56	(0,06)	1,44	(0,04)
KOR	0,60	(0,05)	1,23	(0,04)	0,53	(0,06)	1,35	(0,05)	0,61	(0,05)	1,33	(0,05)	0,48	(0,06)	1,33	(0,04)
LUX	0,43	(0,06)	1,42	(0,04)	0,04	(0,05)	1,49	(0,04)	0,33	(0,05)	1,54	(0,04)	0,05	(0,06)	1,53	(0,04)
LVA	0,62	(0,07)	1,30	(0,06)	0,31	(0,06)	1,36	(0,05)	0,46	(0,06)	1,30	(0,05)	0,22	(0,05)	1,25	(0,04)
NLD	0,83	(0,05)	1,26	(0,04)	0,70	(0,05)	1,32	(0,04)	0,72	(0,05)	1,37	(0,04)	0,53	(0,07)	1,46	(0,04)
NOR	1,04	(0,05)	1,28	(0,04)	0,50	(0,05)	1,48	(0,04)	0,64	(0,06)	1,44	(0,04)	0,46	(0,06)	1,52	(0,04)
NZL	0,73	(0,05)	1,35	(0,05)	0,21	(0,07)	1,60	(0,05)	0,51	(0,06)	1,43	(0,04)	0,26	(0,07)	1,54	(0,06)
POL	0,53	(0,05)	1,22	(0,03)	0,43	(0,06)	1,36	(0,04)	0,47	(0,05)	1,29	(0,04)	0,42	(0,05)	1,30	(0,04)
PRT	0,67	(0,06)	1,33	(0,04)	0,02	(0,06)	1,41	(0,03)	0,42	(0,05)	1,29	(0,03)	0,14	(0,06)	1,42	(0,04)
SVK	0,27	(0,05)	1,41	(0,04)	-0,16	(0,06)	1,38	(0,05)	0,17	(0,05)	1,41	(0,04)	-0,20	(0,06)	1,44	(0,04)
SVN	0,67	(0,06)	1,29	(0,05)	0,55	(0,06)	1,36	(0,04)	0,45	(0,06)	1,38	(0,05)	0,32	(0,06)	1,41	(0,04)
SWE	0,82	(0,06)	1,37	(0,06)	0,20	(0,06)	1,46	(0,04)	0,38	(0,07)	1,56	(0,04)	0,15	(0,06)	1,52	(0,05)
TUR	-0,06	(0,07)	1,15	(0,04)	-0,30	(0,07)	1,16	(0,05)	-0,25	(0,06)	1,18	(0,05)	-0,43	(0,07)	1,31	(0,05)
USA	0,69	(0,06)	1,31	(0,04)	0,21	(0,07)	1,43	(0,04)	0,39	(0,06)	1,36	(0,04)	0,16	(0,06)	1,45	(0,05)

Note. Les données présentées entre parenthèses sont les erreurs-types.

## Annexe 3

*Moyennes et écarts-types de la performance des élèves, par pays et par position, en mathématiques au cycle 2012*

Pays	Position 1				Position 2				Position 3				Position 4			
	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.
AUS	0,10	(0,03)	1,64	(0,02)	0,14	(0,03)	1,68	(0,02)	-0,03	(0,03)	1,70	(0,02)	-0,14	(0,03)	1,74	(0,02)
AUT	0,18	(0,04)	1,62	(0,03)	0,11	(0,04)	1,65	(0,03)	0,04	(0,04)	1,62	(0,03)	-0,08	(0,05)	1,68	(0,03)
BEL	0,35	(0,03)	1,72	(0,02)	0,31	(0,03)	1,71	(0,02)	0,16	(0,04)	1,76	(0,02)	0,01	(0,03)	1,86	(0,03)
CAN	0,33	(0,03)	1,62	(0,02)	0,30	(0,03)	1,64	(0,02)	0,16	(0,03)	1,66	(0,02)	-0,08	(0,03)	1,77	(0,02)
CHE	0,45	(0,05)	1,64	(0,02)	0,44	(0,04)	1,66	(0,03)	0,34	(0,04)	1,68	(0,03)	0,15	(0,05)	1,78	(0,02)
CZE	0,07	(0,04)	1,62	(0,03)	0,03	(0,05)	1,59	(0,03)	0,01	(0,04)	1,61	(0,03)	-0,20	(0,05)	1,69	(0,03)
DEU	0,26	(0,05)	1,65	(0,03)	0,27	(0,05)	1,66	(0,02)	0,18	(0,05)	1,66	(0,03)	-0,02	(0,05)	1,69	(0,03)
DNK	0,08	(0,04)	1,53	(0,03)	0,12	(0,04)	1,55	(0,03)	-0,07	(0,04)	1,61	(0,03)	-0,23	(0,04)	1,65	(0,03)
ESP	-0,11	(0,03)	1,57	(0,02)	-0,17	(0,03)	1,57	(0,02)	-0,34	(0,03)	1,57	(0,02)	-0,61	(0,04)	1,69	(0,02)
EST	0,33	(0,04)	1,54	(0,03)	0,24	(0,04)	1,45	(0,02)	0,26	(0,04)	1,50	(0,03)	0,05	(0,04)	1,58	(0,03)
FIN	0,32	(0,03)	1,53	(0,02)	0,24	(0,03)	1,53	(0,02)	0,20	(0,04)	1,54	(0,02)	0,01	(0,03)	1,66	(0,02)
FRA	0,06	(0,04)	1,60	(0,02)	0,02	(0,04)	1,69	(0,03)	-0,16	(0,05)	1,72	(0,03)	-0,32	(0,05)	1,94	(0,03)
GBR	-0,03	(0,05)	1,64	(0,03)	-0,04	(0,05)	1,70	(0,04)	-0,22	(0,05)	1,72	(0,03)	-0,42	(0,06)	1,75	(0,03)
GRC	-0,46	(0,04)	1,53	(0,02)	-0,57	(0,04)	1,56	(0,02)	-0,88	(0,05)	1,58	(0,03)	-1,20	(0,04)	1,59	(0,03)
HUN	-0,27	(0,05)	1,59	(0,03)	-0,26	(0,05)	1,60	(0,03)	-0,42	(0,05)	1,66	(0,04)	-0,58	(0,06)	1,66	(0,05)
IRL	0,04	(0,03)	1,53	(0,02)	0,06	(0,04)	1,55	(0,02)	-0,05	(0,05)	1,60	(0,03)	-0,15	(0,04)	1,62	(0,03)
ISL	0,07	(0,05)	1,66	(0,03)	-0,07	(0,04)	1,65	(0,03)	-0,26	(0,05)	1,70	(0,03)	-0,50	(0,04)	1,70	(0,03)
ISR	-0,34	(0,07)	1,68	(0,03)	-0,37	(0,08)	1,77	(0,03)	-0,52	(0,06)	1,82	(0,02)	-0,97	(0,08)	1,91	(0,03)

Pays	Position 1			Position 2			Position 3			Position 4						
	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.
ITA	-0,06	(0,03)	1,61	(0,02)	-0,14	(0,03)	1,63	(0,02)	-0,28	(0,03)	1,66	(0,02)	-0,47	(0,03)	1,76	(0,02)
JPN	0,54	(0,05)	1,63	(0,03)	0,47	(0,05)	1,64	(0,03)	0,24	(0,06)	1,73	(0,03)	0,17	(0,06)	1,81	(0,03)
KOR	0,73	(0,06)	1,66	(0,03)	0,66	(0,07)	1,72	(0,02)	0,50	(0,07)	1,74	(0,03)	0,45	(0,07)	1,87	(0,03)
LUX	-0,05	(0,03)	1,63	(0,03)	-0,04	(0,03)	1,66	(0,03)	-0,25	(0,03)	1,72	(0,03)	-0,47	(0,03)	1,80	(0,03)
LVA	-0,18	(0,05)	1,50	(0,03)	-0,24	(0,05)	1,49	(0,03)	-0,24	(0,05)	1,54	(0,03)	-0,38	(0,05)	1,56	(0,03)
NLD	0,46	(0,06)	1,60	(0,03)	0,42	(0,06)	1,66	(0,03)	0,21	(0,05)	1,66	(0,04)	0,07	(0,06)	1,71	(0,04)
NOR	-0,02	(0,04)	1,58	(0,03)	-0,08	(0,04)	1,68	(0,03)	-0,29	(0,05)	1,69	(0,03)	-0,53	(0,06)	1,76	(0,03)
NZL	-0,01	(0,04)	1,63	(0,02)	-0,01	(0,04)	1,72	(0,03)	-0,13	(0,04)	1,75	(0,02)	-0,27	(0,04)	1,77	(0,03)
POL	0,25	(0,06)	1,68	(0,03)	0,23	(0,05)	1,63	(0,03)	0,08	(0,06)	1,68	(0,03)	-0,11	(0,05)	1,69	(0,03)
PRT	-0,01	(0,05)	1,68	(0,03)	-0,10	(0,05)	1,61	(0,03)	-0,25	(0,06)	1,71	(0,03)	-0,61	(0,06)	1,73	(0,03)
SVK	-0,11	(0,05)	1,72	(0,03)	-0,26	(0,05)	1,69	(0,03)	-0,28	(0,05)	1,83	(0,04)	-0,50	(0,06)	1,81	(0,04)
SVN	0,16	(0,04)	1,62	(0,04)	0,10	(0,04)	1,62	(0,03)	-0,02	(0,04)	1,70	(0,03)	-0,22	(0,04)	1,74	(0,03)
SWE	-0,08	(0,04)	1,57	(0,03)	-0,27	(0,04)	1,63	(0,03)	-0,41	(0,04)	1,71	(0,03)	-0,74	(0,05)	1,77	(0,03)
TUR	-0,79	(0,07)	1,62	(0,04)	-0,78	(0,08)	1,59	(0,05)	-0,92	(0,06)	1,60	(0,04)	-1,03	(0,08)	1,65	(0,05)
USA	-0,33	(0,05)	1,55	(0,03)	-0,27	(0,05)	1,58	(0,03)	-0,44	(0,06)	1,59	(0,02)	-0,51	(0,06)	1,69	(0,03)

Note. Les données présentées entre parenthèses sont les erreurs-types.

## Annexe 4

*Moyennes et écarts-types de la performance des élèves, par pays et par position, en mathématiques au cycle 2015*

Pays	Position 1				Position 2				Position 3				Position 4			
	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.
AUS	-0,05	(0,04)	1,59	(0,03)	-0,19	(0,04)	1,66	(0,03)	-0,18	(0,04)	1,72	(0,04)	-0,35	(0,04)	1,67	(0,03)
AUT	-0,15	(0,05)	1,50	(0,04)	-0,15	(0,05)	1,57	(0,04)	-0,24	(0,06)	1,67	(0,04)	-0,28	(0,06)	1,65	(0,04)
BEL	0,14	(0,05)	1,61	(0,03)	-0,01	(0,04)	1,74	(0,03)	-0,04	(0,05)	1,68	(0,03)	-0,12	(0,05)	1,73	(0,03)
CAN	0,13	(0,04)	1,56	(0,02)	-0,03	(0,04)	1,64	(0,02)	0,10	(0,04)	1,62	(0,03)	0,08	(0,05)	1,62	(0,03)
CHE	0,24	(0,06)	1,61	(0,03)	0,04	(0,05)	1,68	(0,04)	0,05	(0,06)	1,67	(0,04)	-0,09	(0,06)	1,76	(0,03)
CZE	-0,06	(0,05)	1,59	(0,04)	-0,15	(0,05)	1,53	(0,04)	-0,20	(0,05)	1,62	(0,04)	-0,30	(0,06)	1,60	(0,03)
DEU	0,04	(0,05)	1,56	(0,04)	-0,04	(0,05)	1,56	(0,04)	-0,04	(0,05)	1,51	(0,04)	-0,04	(0,06)	1,65	(0,04)
DNK	0,24	(0,06)	1,52	(0,05)	0,02	(0,07)	1,54	(0,03)	-0,01	(0,06)	1,56	(0,04)	-0,01	(0,06)	1,61	(0,04)
ESP	-0,28	(0,06)	1,58	(0,03)	-0,45	(0,05)	1,59	(0,04)	-0,37	(0,05)	1,54	(0,03)	-0,49	(0,05)	1,46	(0,03)
EST	0,23	(0,05)	1,52	(0,04)	0,13	(0,05)	1,53	(0,03)	0,10	(0,06)	1,55	(0,04)	0,00	(0,05)	1,58	(0,03)
FIN	0,01	(0,05)	1,45	(0,04)	-0,10	(0,05)	1,42	(0,04)	0,01	(0,04)	1,47	(0,04)	-0,10	(0,05)	1,46	(0,04)
FRA	-0,12	(0,04)	1,61	(0,03)	-0,19	(0,05)	1,69	(0,04)	-0,20	(0,06)	1,71	(0,03)	-0,37	(0,05)	1,79	(0,04)
GBR	-0,09	(0,05)	1,62	(0,04)	-0,16	(0,06)	1,74	(0,04)	-0,22	(0,05)	1,64	(0,03)	-0,35	(0,05)	1,67	(0,03)
GRC	-0,59	(0,06)	1,52	(0,03)	-0,76	(0,05)	1,50	(0,04)	-0,75	(0,08)	1,56	(0,04)	-0,95	(0,06)	1,47	(0,04)
HUN	-0,36	(0,05)	1,58	(0,04)	-0,48	(0,06)	1,59	(0,04)	-0,43	(0,05)	1,59	(0,05)	-0,55	(0,06)	1,65	(0,04)
IRL	-0,08	(0,05)	1,60	(0,03)	-0,08	(0,05)	1,55	(0,03)	-0,13	(0,04)	1,54	(0,03)	-0,08	(0,04)	1,63	(0,03)
ISL	-0,24	(0,06)	1,62	(0,05)	-0,50	(0,05)	1,50	(0,05)	-0,31	(0,07)	1,66	(0,06)	-0,47	(0,07)	1,63	(0,05)
ISR	-0,34	(0,06)	1,72	(0,04)	-0,68	(0,06)	1,74	(0,03)	-0,57	(0,07)	1,72	(0,04)	-0,77	(0,07)	1,79	(0,03)

Pays	Position 1			Position 2			Position 3			Position 4						
	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.	$\mu$	E. T.	$\sigma$	E. T.
ITA	-0,16	(0,05)	1,54	(0,04)	-0,41	(0,05)	1,54	(0,04)	-0,27	(0,06)	1,64	(0,04)	-0,44	(0,06)	1,66	(0,04)
JPN	0,21	(0,05)	1,61	(0,03)	0,22	(0,05)	1,67	(0,03)	0,26	(0,06)	1,66	(0,04)	0,19	(0,06)	1,66	(0,03)
KOR	0,29	(0,06)	1,67	(0,04)	0,28	(0,06)	1,67	(0,04)	0,15	(0,07)	1,76	(0,04)	0,15	(0,07)	1,77	(0,04)
LUX	-0,25	(0,05)	1,63	(0,04)	-0,40	(0,06)	1,67	(0,04)	-0,38	(0,05)	1,66	(0,04)	-0,38	(0,05)	1,65	(0,04)
LVA	-0,43	(0,05)	1,34	(0,04)	-0,53	(0,06)	1,50	(0,04)	-0,56	(0,05)	1,38	(0,04)	-0,52	(0,05)	1,41	(0,03)
NLD	0,22	(0,05)	1,62	(0,03)	0,19	(0,05)	1,66	(0,04)	0,01	(0,04)	1,59	(0,04)	-0,12	(0,05)	1,69	(0,04)
NOR	-0,08	(0,05)	1,59	(0,03)	-0,22	(0,05)	1,58	(0,04)	-0,21	(0,06)	1,58	(0,04)	-0,38	(0,06)	1,64	(0,04)
NZL	-0,06	(0,06)	1,58	(0,04)	-0,28	(0,08)	1,72	(0,04)	-0,19	(0,07)	1,64	(0,04)	-0,28	(0,06)	1,66	(0,04)
POL	-0,09	(0,06)	1,70	(0,04)	-0,15	(0,06)	1,62	(0,04)	-0,08	(0,06)	1,65	(0,04)	-0,12	(0,06)	1,67	(0,04)
PRT	-0,11	(0,07)	1,73	(0,04)	-0,47	(0,06)	1,57	(0,04)	-0,24	(0,06)	1,69	(0,04)	-0,39	(0,06)	1,71	(0,04)
SVK	-0,26	(0,05)	1,62	(0,04)	-0,60	(0,05)	1,54	(0,03)	-0,54	(0,05)	1,61	(0,04)	-0,57	(0,06)	1,59	(0,04)
SVN	0,12	(0,06)	1,50	(0,04)	0,04	(0,06)	1,57	(0,04)	-0,07	(0,06)	1,63	(0,05)	-0,04	(0,06)	1,63	(0,05)
SWE	-0,08	(0,06)	1,60	(0,03)	-0,39	(0,08)	1,64	(0,05)	-0,33	(0,06)	1,64	(0,04)	-0,59	(0,06)	1,63	(0,04)
TUR	-1,10	(0,08)	1,54	(0,05)	-1,15	(0,07)	1,52	(0,04)	-1,23	(0,07)	1,49	(0,04)	-1,23	(0,07)	1,50	(0,03)
USA	-0,42	(0,06)	1,61	(0,04)	-0,58	(0,06)	1,56	(0,05)	-0,54	(0,06)	1,63	(0,04)	-0,65	(0,07)	1,61	(0,03)

Note. Les données présentées entre parenthèses sont les erreurs-types.