The learning gain over one school year among 15-year-olds: An international comparison based on PISA

Francesco Avvisati, Pauline Givord

SciencesPo
LABORATOIRE INTERDISCIPLINAIRE
D'ÉVALUATION DES POLITIQUES PUBLIQUES

LIEPP Working Paper
July 2022, n°136

The learning gain over one school year among 15-year-olds: An international comparison based on PISA

Francesco AVVISATI
OECD
francesco.avvisati@gmail.com

Pauline GIVORD
INSEE - LIEPP
pauline.givord@insee.fr

www.sciencespo.fr/liepp

Comment citer cette publication :
AVVISATI, Francesco, GIVORD, Pauline, **The learning gain over one school year among 15-year-olds: An international comparison based on PISA,** *Sciences Po LIEPP Working Paper* n°136, 2022-07-07.

The learning gain over one school year among 15-year-olds: An international comparison based on PISA*

Francesco Avvisati†
Pauline Givord‡

July 5, 2022

Abstract

We compare the learning gain over one year of schooling among 15-year-old students in Austria, Brazil, Malaysia, Scotland (United Kingdom) and Singapore. Common metrics for reading, mathematics and science learning, as established by the Programme for International Student Assessment (PISA), are used. In order to overcome the limitations of a cross-sectional, single-cohort design, we combine multiple vintages of PISA data and exploit the fact that the testing period in these countries varied over the years. The results show that students' yearly learning progress around the age of 15 varies from about one-tenth of a standard deviation in students' test scores in Malaysia to about one-fourth of a standard deviation or more in Austria and Scotland.

JEL classification: I21, I25, I26

1 Introduction

How does the pace of learning – i.e. the gain in knowledge and skills associated with one grade of schooling, or *grade gain* – compare across countries? International assessments are designed to compare learning outcomes at a particular point in students' school career, but they do not directly show how the learning gains made by students over comparable time intervals differ across countries. Some international assessments, such as the Trends in International Mathematics and Science Study (TIMSS), assess students who are in a particular grade level (typically, 4^{th} and 8^{th} grade); however, because of grade retention or grade skipping, in many countries students who are in the same grade level may have been exposed to different amounts of schooling. More generally, students start school with widely different knowledge and skills, meaning that differences in mean scores in these assessments cannot be readily interpreted as a measure of school productivity. The Programme for International Student Assessment (PISA), in contrast, assesses the learning of students at age 15, regardless of the grade level attended and the amount of schooling to which students have been exposed; its mean scores cannot indicate school productivity without strong assumptions.

*The authors would like to thank Nathan Viltard for his research assistance and contribution to this paper during his internship with the OECD Directorate for Education and Skills. We are also indebted to those who participated in the Applied Economics Work-in-Progress Seminar at the OECD (October 2020) and the Applied Economics Lunch Seminar at the Paris School of Economics. Previous versions of this manuscript benefited from valuable feedback from Tiago Fragoso, Marco Paccagnella and Andreas Schleicher, as well as from experts at the Federal Institute for Quality Assurance in the Austrian School System (IQS). The authors also thank Julie Harris for editing a draft version of this article. The opinions expressed and arguments employed in this article are those of the authors and do not necessarily represent the official view of the Insee, the OECD or of its member countries. Francesco Avvisati gratefully acknowledges support from the French National Research Agency (ANR-21-ESRE-0034)

†OECD, 2 rue André Pascal, 75775 Paris Cedex 16, France. Corresponding author, present address: J-PAL Europe, 48 Boulevard Jourdan, 75014 Paris, France; francesco.avvisati@gmail.com

‡Insee-LIEPP

This article provides estimates of the yearly average learning gain in five jurisdictions that span a wide range of levels of economic development across three continents. Despite the large differences among these five school systems, the fact that results are expressed in terms of common metrics makes it possible to compare the relative effectiveness of learning systems around the age of 15 years, in three subjects (reading, mathematics and science).

The results show that students' yearly learning progress around the age of 15 varies from about one-tenth of a standard deviation in students' test scores in Malaysia to about one-fourth of a standard deviation or more in Austria and Scotland (United Kingdom), with the estimates for Brazil and Singapore comprised in-between these extremes. Results also show that in general, the average pace of learning of 15-year-olds does not differ significantly between boys and girls, or between advantaged students and disadvantaged students (defined by a median split on a variable measuring the social, economic and cultural status of students' households). A significant difference in the yearly learning gain by subgroup is observed only in Scotland, where boys appear to make stronger progress than girls around the age of 15. This suggests that in all five countries, the wide socio-economic gaps observed at age 15, as well as the gaps in reading performance in favour of girls, largely reflect skill differences that appear at a younger age.

Several articles, in the past, have tried to identify average learning progressions from student assessment data, but only a few have done so in ways that allow to compare such progressions across countries. At first sight, the main obstacle in this endeavour is the fact that longitudinal data collections at international level are extremely rare. Only a few prior studies have embedded internationally linked assessments in longitudinal data collections (Singh 2019, Jones, et al. 2014, Prenzel, et al. 2006, Nagy, et al. 2017). Singh (2019) highlights significant differences in productivity across developing countries, and shows that in primary grades, the effect of an extra grade of schooling on test scores is substantially higher in Viet Nam than in Andra Pradesh (India), Ethiopia or Peru. Prenzel et al. (2006) and Nagy et al. (2017) provide estimates for 15-year-olds in Germany, based on the PISA test.[1]

Despite its intuitive appeal, the identification of average learning gains from longitudinal data (i.e., from the aggregation of individual learning gains) faces two major challenges. A first challenge is to ensure that individual learning-gain estimates are not confounded by variations in students' motivation or in other contingent factors that are unrelated to students' learning. A second challenge, for the inference to a broader population of interest, is the possibility of selective attrition: due to school drop-out, absenteeism, grade repetition or for other reasons, the student population for which both baseline and follow-up test scores are available may differ in significant

[1]The underlying assumption when computing averages and when comparing test-score differences over time, across groups, or between countries, is that test scores are reported on an interval scale, meaning that at any point in the distribution of students' scores, a one unit change reflects the same change in the underlying knowledge or skill. In essence, however, the measurement scale of test scores is ordinal (Bond and Lang, 2013; Schroeder and Yitzakhi, 2020).
Jacob and Rothstein (2017) discuss the issue in detail: the "interval" property of test scores appears particularly arbitrary when test scores simply reflect the fraction of correct answers to a particular set of questions. In contrast, PISA scores rely on principled procedures for test design and scaling, which ensure that different sets of items measuring the same skill produce not only identical rankings of individual test takers, but also the same ordering of test-score differences across pairs of test takers (for a discussion see for instance Braun and von Davier, 2018). In particular, PISA scores are derived from parametric item-response-theory models, which link students answers to the test to a latent ability parameter (the score). The weight assigned to different component traits in the underlying test is the result of a consensus among participating countries and of item-selection procedures which are informed by the model's hypotheses, so as to retain only those questions whose responses show adequate fit with the model's predictions. As a result of the model's parametric assumptions, within each country, PISA scores are close to normally distributed.
Jacob and Rothstein (2017) also raise concerns about the use, in econometric analyses, of plausible values (i.e. multiply imputed test scores, provided in public-use files for secondary analyses). In order to result in unbiased results, the "conditioning model" used for the generation of plausible values must indeed incorporate all variables included in the regression analysis. While Jacob and Rothstein suggest that this may not always be the case (in particular, when interest lies in combining test data with external information), all variables that are used in our analysis, and in particular, the students' month of birth, are included in the conditioning model used to generate individual students' PISA scores (see e.g. OECD 2017, Annex B).
Examples of recent studies using PISA scores – and treating them as interval – include Lavy (2015), Hanushek et al. (2019), and De Philippis and Rossi (2020).

ways from the population of interest.

In the absence of longitudinal data, a regression-discontinuity estimator may identify yearly learning progressions in international assessment data, and may be more robust to the above challenges (contextual and selection effects). We explore this approach in a companion paper (Avvisati and Givord, 2021). Our estimator compares students born "just before" and "immediately after" the cut-off date for first-grade enrolment, in countries where the population assessed in international tests did not coincide with the school-entry cohort. In contrast to prior studies that rely on a similar identification strategy, our companion paper suggests focusing on comparing, in these countries, the eldest and the youngest students in terms of age (rather than in terms of school-entry age). Indeed, because at any given point in time, students' age, their length of schooling, and their age at school entry are bound by a simple additive relationship, it is not possible to identify their distinct contribution to test scores in a cross-sectional design. While prior studies compared students of (almost) same age to focus on the joint effect of schooling and of school-entry age (sometimes claiming, for the interpretation of their results, that school-entry age effects are negligible),[2] our companion paper compares students with the same school-entry age to focus on the joint effect of schooling and age; the latter corresponds to the quantity identified in studies based on longitudinal data. Due to the limited sample size in survey datasets, however, cross-country comparisons of estimates based on a regression-discontinuity estimator are often imprecise.

This study proposes a different identification strategy to estimate this joint effect of age and length of schooling on learning. We use the change, across different vintages of PISA, in the time of the year when the test was conducted. In PISA, the birth dates of eligible students depend on the testing date; and when the testing date changes, the month of birth of the eldest eligible students also changes. Thus, when grouping students by month of birth, two groups can be defined such that the change in testing date has opposite effects on their age and length of schooling at the time of the test: students born in certain months are assessed at a younger age and at an earlier point in their school career than would have been the case, had the testing date remained the same; in contrast, students born in the remaining months are assessed at an older age and at the beginning of the following grade. The change in testing date thus acts as an exogenous source of variation which allows for the identification of the full effect of a year of schooling and of age through a difference-in-difference estimator.

2 PISA data, testing dates and samples

All data used in this article were collected by the Programme for International Student Assessment (PISA), a large-scale, cross-national assessment of the reading, mathematics, and science performance of 15-year-old students. PISA has been administered to samples of 15-year-old students across almost 100 countries in total, every three years since 2000 (participation of countries has generally increased over time, but not all countries participated in every assessment cycle since they began taking part in PISA). Results for all three domains of reading, mathematics and science are fully comparable over time starting with PISA 2006.[3]

PISA standards, which apply to all countries and economies participating in PISA, specify that "Unless otherwise agreed upon, the testing period [...] begins exactly three years from the beginning of the testing period in the previous PISA cycle" (Standard 1.3). This consistency in testing dates ensures the comparability over time of results, which may otherwise be influenced by

[2] The effects of students' age at school entry on learning have been the focus of much attention in the economics of education literature (Dearden, Crawford and Meghir 2010, Black, Devereux and Salvanes 2011, Bedard and Dhuey 2006). Givord (2020) reviews this literature and provides international evidence based on PISA data.

[3] All data used in the present article are available as "public use files" and can be accessed through `www.oecd.org/pisa`. PISA test scores are norm-referenced scales derived from student responses to a test using item-response-theory (IRT) models. For each subject, the test norm was set to a mean of 500 and a standard deviation of 100 across students from OECD countries in a baseline year (which varies by subject), and all later tests have since been reported on the same scale. The Stata package "repest" was used for the main analyses (Avvisati and Keslair 2014).

contextual effects (e.g. seasonal fluctuations in students' motivation to complete a low-stakes test). Occasionally, however, countries request and are permitted to change their testing dates. Over recent cycles (i.e. between 2006 and 2018), this has been the case in five jurisdictions: Austria, Brazil, Malaysia, Scotland (United Kingdom) and Singapore.

In all five jurisdictions, the decision to change the testing period was driven mainly by logistic considerations. Austria joined the PISA 2015 cycle late and was therefore allowed to move its testing period to the end of the calendar year; in 2018, the testing dates returned to a period around April (as in earlier cycles). Brazil changed its testing date in 2009, from July/August to April/May, in order to simplify the process of compiling the lists of eligible students (Gomes, Hirata and Oliveira 2020): after this change, the eligible students were all those born in a particular calendar year. Malaysia and Singapore participated in PISA for the first time in 2010 and 2009; they changed their testing date in 2012, for their second participation in PISA, most likely in order to avoid issues encountered in their first participation in PISA (such as the need to assess students in different grade levels: the new testing date made the eligible students coincide with a single school-entry cohort). Scotland used to administer PISA towards the beginning of the calendar year (and towards the end of the school year), unlike the rest of the United Kingdom; in 2018, Scotland moved its testing dates to the Northern Hemisphere fall (October/November), aligning them more closely with those of the rest of the United Kingdom. By moving the PISA test to the fall, the Scottish authorities in charge of PISA administration intended to increase the comparability of PISA results with the results of students in England, Wales and Northern Ireland, and at the same time reduce the pressure on schools and students at the end of the school year, which coincides with an exam period.

A change in testing dates automatically results in a change in the birth dates of eligible students in PISA. In both Austria and Scotland, for example, the PISA cohort comprised all students born in a particular calendar year when the test was conducted in spring (towards the end of the school year). The eldest eligible students were those born in January; and the youngest students were those born in December. However, in Austria in 2015, and in Scotland in 2018, when the test was conducted in autumn, the PISA cohort spanned two calendar years, and the eldest eligible students were those born in August of the first year.

The actual grade level of students participating in PISA depends mainly on their month of birth (unless the testing period is chosen so that the PISA cohort coincides with a school-entry cohort). Indeed, in most jurisdictions (including the five examined in this article), school-entry regulations are centred around a cut-off date that determines eligibility for enrolment in first grade, and define the birth date of the eldest children in consecutive school-entry cohorts. School-entry regulations in Austria, for example,[4] define the school-entry cohort as the cohort of children who turned six between 1 September of the previous year and 31 August of the current year. Based on school-entry regulations alone, one would therefore expect that, among those born in the same calendar year, students born between January and August in Austria have attended school for one year less than the remaining students. In practice, the actual grade of students at age 15 can deviate from the expected grade because of deferred entry, grade repetition or other circumstances. A simple plot of the actual grade observed in PISA by month of birth shows, however, that the theoretical grade is a strong predictor of the actual grade (Figure 1). It also shows how this month-to-grade mapping depends on the testing date, which determines the birth date of the youngest and eldest students eligible to sit the PISA test.

Table A.1 in the Appendix presents descriptive statistics for the samples and main variables of interest used in this paper. As expected, given the PISA sampling design, the average age of students included in PISA varied little across countries and years and remained between 15.7 and 15.9 years, and the gender composition remained broadly stable. Similarly, the average number of completed grade levels remained stable, across the full cohort of participating students, even when the testing date changed. The only significant change in the average number of grade levels completed over the period is observed in Brazil, where the grade level of the cohort assessed in 2012 and in later years reflects the federal school reform of 2006, which lowered the school starting

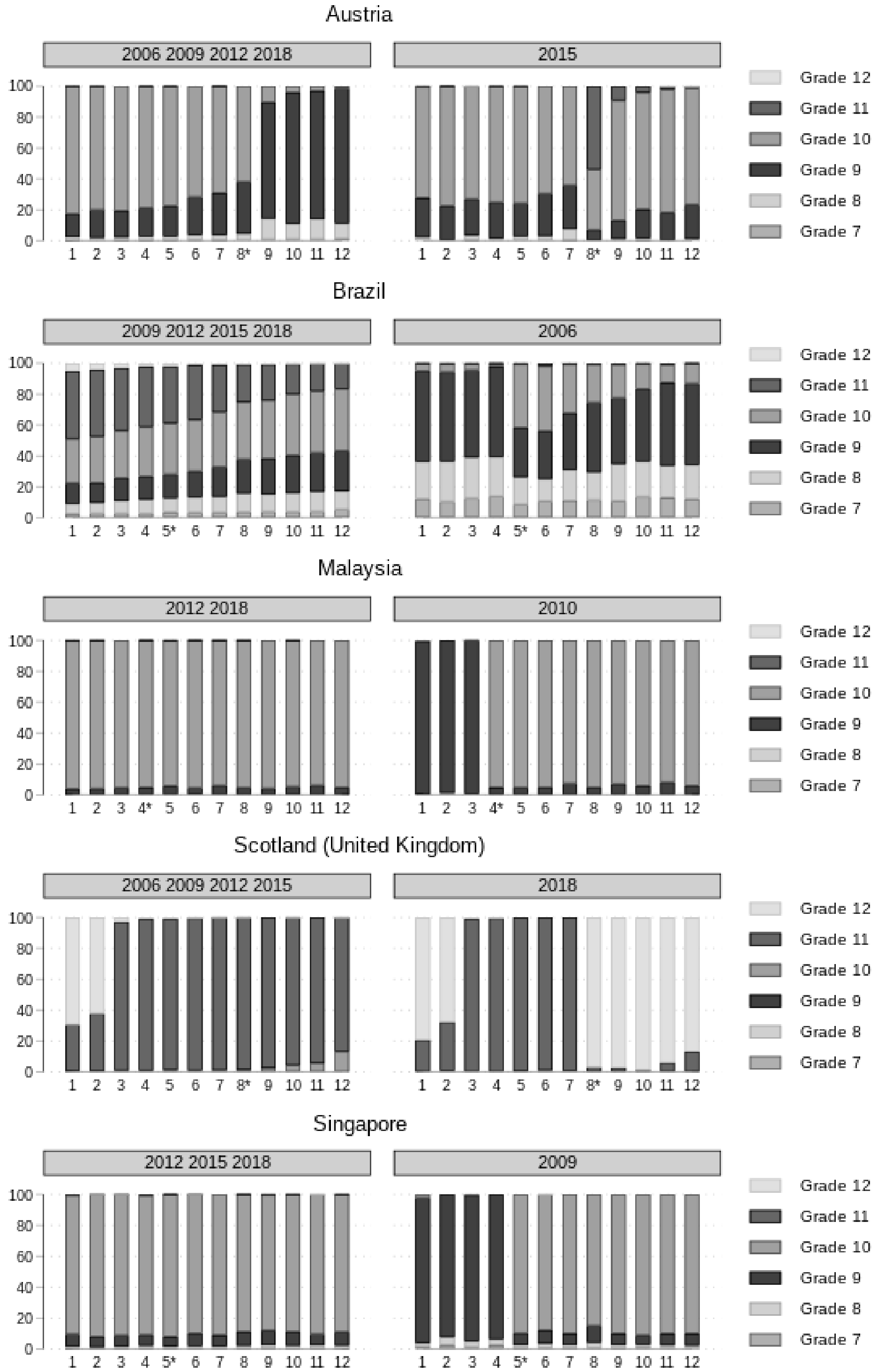
[4]Throughout the article, to simplify exposition, we will use Austria as an illustrative example.

age.[5] Other characteristics of the sample changed to a greater extent, including mean performance of students and their family background. Throughout the period, student performance is highest in Singapore, followed by Austria and Scotland; Malaysia and Brazil scored below these countries. In 2018, the difference in mean performance between Singapore and Brazil was about 190 score points in mathematics, 140 score points in reading and 150 score points in science. The proportion of students with an immigrant background increased over time in Austria, Malaysia, Scotland and Singapore. The proportion of students who reported one of their parents to have a tertiary degree increased, in particular, in Malaysia, Scotland and Singapore, reflecting the expansion of access to tertiary education among earlier generations. This proportion decreased at first in Brazil, most likely because of differences in population coverage across PISA waves:[6] in Brazil, only 55% of all 15-year-olds were covered by PISA in 2006; this proportion increased to 70% by 2012 as a result of the educational expansion and of the greater access of disadvantaged children to secondary schooling (OECD 2019a, Table I.A2.2).

[5]A federal law, enacted in 2006, required all schools to enrol 6-years old students in elementary school by the year 2010 (Law 11.274 of 2006) (Rosa, Martins and Carnoy 2019). The numbering of grade levels in PISA datasets after 2012 reflects this change, even though in practice, only a minority of the students assessed in 2012 really had started school earlier than previous cohorts (if their municipalities had anticipated the reform). For most, the change was likely only nominal. Up to 2018, all PISA cohorts were only partially affected by the reform: students eligible for PISA in 2018, and born in 2002, were expected to start school during the transition period between the approval of the law and the deadline for its enforcement (2010).

[6]PISA samples are representative of students who are enrolled in Grade 7 or above and who are between 15 years and 3 months and 16 years and 2 months at the time of the assessment administration (generally referred to as 15-year-olds in this article). In countries where many 15-year-olds are found in lower grades or out of school, the target population may represent only a fraction of all 15-year-olds.

Figure 1: Grade distribution by month of birth



Notes: Charts on the left refer to early-testing years; charts on the right to late-testing years. In early-testing years, testing occurs towards the beginning of the calendar year (typically, around March-April) and the PISA cohort coincides with all students born in a particular calendar year (with the eldest eligible students born in January). In late-testing years, testing occurs later in the calendar year (at a different date, depending on the country) and the birth dates of PISA-eligible students span across two years; the month of birth of the eldest eligible student in a late-testing year is marked by an asterisk (*).
Source: PISA 2006, 2009, 2012, 2015 and 2018 datasets, `https://www.oecd.org/pisa/data/` (accessed on 17 May 2021).

3 Identification strategy

In a cross-sectional, single-cohort study such as PISA, the observed variation in the number of completed grade levels is endogenous: decisions to anticipate or delay entry into first grade, as well as grade retention and grade skipping, are typically influenced by factors that are difficult to observe (including prior performance, family involvement, etc.), and that may also exert a direct influence on learning outcomes. This endogeneity implies that naïve comparisons of students who are found in different grades do not only reflect the effect of the additional schooling attended by such students but also the many other observed and unobserved differences between these students.

To address this endogeneity issue, the student's month of birth (and the expected number of grade levels completed) may be used as an exogenous source of variation in the actual grade level. Indeed, in most countries, school-entry regulations rely on a cut-off date that determines eligibility for enrolment in first grade, and defines the birth date of the eldest children in consecutive school-entry cohorts.

However, this strategy gives rise to another identification issue. If students are observed only once, the variation in test results around the cut-off date for first-grade enrolment can be interpreted as reflecting a "grade effect" only under strong assumptions about the effect of students' age at school entry (age-at-entry effects). Indeed, such effects, if they exist, cannot be accounted for separately, since the expected age at entry, the expected number of grade levels completed and the current age of the student are linked by a simple, additive relationship ($age_{is} = expgrade_{is} + expentryage_{is}$).

In this paper, these challenges are addressed by exploiting a source of exogenous variation in grade and age that exerts its influence at aggregate levels, when combining multiple PISA samples characterised by some variation in testing dates.

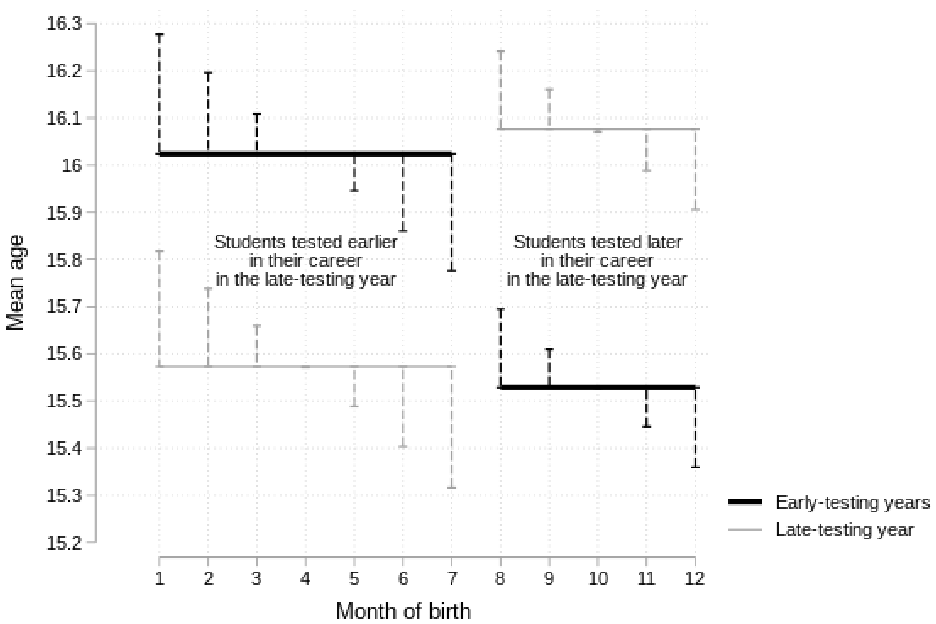
3.1 Identification of grade-and-age effects

The identification strategy to estimate the grade gain relies on comparing, within each education system, the PISA scores of students born in the same calendar month across survey cycles that differ in terms of testing dates. In the case of Austria, the testing dates observed in PISA begin either in March or October; survey cycles are referred to as "early-testing years" when testing began in March and as "late-testing years" when testing began in October. In Brazil, testing began in July in 2006 ("late-testing year") and in March in all following years ("early-testing years"). Similarly, in Malaysia and Singapore, testing began in June or July in their first year of participation (2009/2010) and in March in all following years. Because only students born within a particular 12-month window are eligible to participate in PISA, the testing dates determine the age at which students born in a particular month participate in the PISA test. For example, students eligible to participate in PISA who are born in May are expected to be 15 years and 9 months old if they sit the PISA test at the beginning of March, but only 15 years and 4 months old if they sit the PISA test at the beginning of October. Together with school-entry regulations, testing dates also determine the expected amount of school years completed by students born in a particular month. For example, if students born in May are expected to enter first grade in September at the age of 6 years and 3 months, they will have completed 9 years and 6 months of schooling if they participate in a PISA survey conducted in March, but only 9 years and 1 month of schooling if they participate in a PISA survey conducted in October. As this example shows, when testing dates change and in the absence of changes to school-entry regulations, age at testing and the expected amount of schooling shift in the same direction, and by the same number of months, for students with the same birthday. Each comparison by month of birth across early- and late-testing years thus reflects, among other factors, a particular difference in students' age and amount of schooling.

The key observation for the identification of grade-and-age effects is that grade-and-age differences between late- and early-testing years are negative for some birth dates (for which eligibility criteria imply that participating students are younger by n months when testing is conducted later in the year, as is the case for students born in May in the previous example); but positive for other birth dates (those comprised between August and December, in the case of Austria). Indeed, the age-based definition of eligibility adopted by PISA implies that the *average* age of students in the PISA sample does not change when the date of testing shifts. As a result, by combining the negative grade-and-age shift for students born in certain months with the positive shift for students born in the remaining months, it is possible to observe, indirectly, a difference of a full year of age and a full grade.

Figure 2 presents the graphical intuition behind the difference-in-difference estimator, using Austria as an example. It compares the average age of students in early- and late-testing years, with students divided into two groups depending on their month of birth. The first group includes all students born between January and July (the cut-off month which defines a school-entry cohort). By virtue of the age-based eligibility criteria for PISA, students born in these months were less advanced in their school career (and younger, by a few months) in the year in which testing was conducted later than in the years when it was conducted earlier. The second group includes students born between August and December, who were more advanced in their school career (and older, by a few months) in the year in which testing was conducted later. For students born at the beginning of the calendar year, the difference in age and amount of schooling across late- and early-testing years is of opposite sign than for students born at the end of the calendar year; and the arithmetic sum of these age-differences, taken with the same sign (i.e., the double difference), corresponds to exactly one year of age and of schooling.

Figure 2: Mean age in Austria, by month of birth and year



Notes: Horizontal lines indicate the average age of students born in the corresponding months, by year. Drop lines show the average age for each single month. In Austria, 2006, 2009, 2012 and 2018 were early-testing years, and 2015 was a late-testing year.
Source: PISA 2006, 2009, 2012, 2015 and 2018 datasets, `https://www.oecd.org/pisa/data/` (accessed on 17 May 2021).

Figure 2 implies that the same double-difference estimator, applied to variables other than age and schooling, will reflect the effect of one additional year of age and schooling on these variables. And, under a "common trends" assumption which will be detailed further below, it will reflect *only* the effect of one additional year of age and of schooling.

Differences in performance between early- and late-testing years for students born in a particular month can also reflect a number of other differences beyond this difference in age and (expected) amount of schooling. In particular, there may be differences not only in quantity but also in the quality of education experienced by different cohorts of students: for instance, in Brazil, the cohorts assessed in PISA 2012 or later started primary school at a younger age compared to the cohorts assessed in earlier cycles, as a result of an education reform (Rosa, Martins and Carnoy 2019). Furthermore, the composition of each cohort may differ: for example, Table A.1 in the Appendix shows that in many countries, the share of 15-year-old students whose parents had completed some tertiary education increased over time. There may also be seasonal patterns in test performance or in students' motivation to take a low-stakes test such as PISA.

The essence of the "common-trend" assumption is that seasonal patterns of performance, cohort-specific trends and changes in sample composition are unrelated to a student's month of birth. In the case of Austria, for example, we assume that, on average, the same change in score would have been observed between 2015 (late-testing year) and 2018 (early-testing year) among students born between August and December as among students born between January and July, had a younger cohort of August-to-December born students sat the test in 2015 (the cohort that was one grade below and one year younger than the one that actually sat the test). Under this "common-trend" assumption, a double-difference strategy will net out these confounding factors and can identify the grade-and-age effect.

Formally, let y_{ist} represent the performance in PISA of student i, attending school s, in year t, in a particular country and subject (grade-and-age effects are estimated using separate regressions for each country and subject). Let m_i represent the student's month of birth, and further assume that the performance of student$i$ in PISA can be described by the following additive function:

$$y_{ist} = \alpha_t + \beta^{'} x_{ist} + \sum_{m=1}^{12} \gamma_m \mathbb{1}_{m_i=m} + \delta \mathbb{1}_{t\in L, m_i>M} + \varepsilon_{ist} \tag{1}$$

In this equation, α_t (a year fixed effect) represents the contribution to performance of the average quality of schooling experienced by 15 year olds up to year t and of other factors common to all students in a given year; β' (a vector) represents the influence of student i's characteristics x_{ist} (namely gender, immigrant background and socioeconomic status) on performance; ε_{ist}, an error term, captures the influence of other student- and school-level characteristics on performance, and γ_m captures the effect of a student's month of birth on his or her performance. As discussed earlier, this effect may appear because of at least three main reasons:

1. the student's age on the day of the test, which in PISA varies between 15 years and 3 months and 16 years and 2 months;

2. the amount of schooling received by the student up to the testing date, i.e. the current grade level of the student, minus the fraction of that grade level that remains to be completed; and

3. the student's age at school entry, which can influence (particularly in the early primary grades) children's ability to benefit from schooling and the characteristics of the peer group, and which can, through these two channels, have a lasting influence on students' learning.

Using dummies for the month of birth means that these effects are specified in a flexible way. The common-trend assumption, embedded in Equation 1, implies that the three effects associated with a student's month of birth (age-at-testing or maturity effects, length-of-schooling effects, and age-at-school-entry effects) do not vary over time (other than in ways that are common across all birth dates, captured by α_t). In other words, conditionally on the sample composition (in terms of gender, immigrant background and socio-economic status), all cohort-specific determinants of performance are unrelated to a student's month of birth (i.e. to his or her age, grade level, and expected age at school entry, at least locally, i.e. within the limited range of variation considered).

As illustrated in Figure 2, when the testing date changes, age at testing and length of schooling change in a discontinuous and discrete way across students' birth dates. This is captured by the main parameter of interest δ, which represents the effect of being older by one year and having completed one more year of schooling. It is estimated using the fact that, when testing is conducted later in the year (late-testing years, $t \in L$), PISA measures the performance of students born towards the end of the calendar year ($m_i > M$, where M is the month of birth of the youngest student in PISA in a late-testing year) who are older and more advanced in their schooling, compared to students with the same birth dates who would have been eligible for PISA in March (early-testing years, $t \notin L$). In the equation, the late-testing year is denoted by $t \in L$, where $L = 2015$ in Austria, $L = 2006$ in Brazil, $L = 2010$ in Malaysia, $L = 2018$ in Scotland and $L = 2009$ in Singapore (also see Figure 1). Meanwhile, students born towards the beginning of the calendar year ($m_i \leq M$) are subject to an opposite effect of late-testing years: they are tested at a younger age and earlier in their school career. This effect is captured by the year dummy α_L along with all other year effects, so that δ captures the effect of a full year of age and schooling.

The common-trend assumption may be violated, for example, if among students who sat the test in late testing years, only those born in particular months were touched by an education reform that affected performance, such as a change in grade-repetition practices. The assumption is also violated if differences in unobserved student characteristics across groups defined by the month of birth vary over the years; the influence of such unobserved student characteristics is represented by the error term ε_{ist} in Equation 1. For example, suppose that students born at the end of the calendar year are expected to be in Grade 9 when testing is conducted at the end of the school year, but in Grade 10 when testing is conducted at the beginning of the school year; and further assume that weaker students are likely to drop out of school after Grade 9. As a result, the difference between late- and early-testing years for students born at the end of the calendar year not only reflects the higher age and the greater amount of schooling in the late-testing year but also the selective drop-out of weaker students between Grades 9 and 10; but the latter selection effect is not present (and therefore contributes to δ and is not captured by α_t) for students born at the beginning of the calendar year, who are expected to be in Grade 9 regardless of the testing period. Robustness checks aiming at testing the sensitivity of the main results to these potential violations of the common trend assumptions are presented in Section 5.

3.2 Identification of subgroup differences

In order to explore the existence of differential grade gains depending on students' characteristics, a modified version of Equation 1 is estimated. This version includes additional interaction terms, so that both the underlying trends (represented by α_t in Equation 1), the month-of-birth effects (represented by γ_m) and the grade-gain coefficient δ are allowed to vary across subgroups ($g_i = G$ and $g_i \neq G$). The subgroup indicator $\mathbb{1}_{g_i=G}$ is also included among the vector of control variables x_{ist}).

$$\begin{aligned} y_{ist} =& \alpha_t^0 + \alpha_t^1 \mathbb{1}_{g_i=G} + \beta' x_{ist} + \sum_{m=1}^{12} (\gamma_m^0 \mathbb{1}_{m_i=m} + \gamma_m^1 \mathbb{1}_{m_i=m} \mathbb{1}_{g_i=G}) \\ &+ \delta^0 \mathbb{1}_{t \in L, m_i > M} + \delta^1 \mathbb{1}_{t \in L, m_i > M} \mathbb{1}_{g_i=G} + \varepsilon_{ist} \end{aligned} \tag{2}$$

4 Results

Table 1 shows the average grade-and-age effects estimated by exploiting the variation in testing dates between consecutive PISA cycles (estimates based on Equation 1) for all five countries. The first set of estimates, in the top panel of Table 1, does not include any control variables; a simple graphical representation, similar to Figure 2, illustrates the intuition behind these estimates.

Figure 3 compares, using Austria as an example, the average performance of students in early- and late-testing years, with students divided into two groups depending on their month of birth (the same groups as in Figure 2). In the absence of an overall improvement or decline in performance in the late-testing year, one would expect that students born in the same month, but who are

Table 1: Grade-and-age effects

Country	Nb of obs.	Maths	Reading	Science	Year fixed effects	Month-of-birth dummies
No controls						
Austria	30080	26.38*** (3.25)	31.48*** (3.40)	26.26*** (3.21)	Yes	Yes
Brazil	82458	10.56*** (3.50)	14.61*** (3.89)	11.04*** (3.27)	Yes	Yes
Malaysia	16307	11.38*** (3.58)	8.70** (3.82)	6.55* (3.38)	Yes	Yes
Scotland (UK)	14129	32.91*** (6.34)	34.26*** (5.07)	26.46*** (6.10)	Yes	Yes
Singapore	23620	23.10*** (3.16)	15.85*** (2.98)	18.53*** (3.17)	Yes	Yes
With controls for covariates[1]						
Austria	29422	24.74*** (3.13)	29.40*** (3.19)	24.62*** (2.95)	Yes	Yes
Brazil	77559	11.65*** (3.38)	14.87*** (3.59)	11.76*** (3.05)	Yes	Yes
Malaysia	15921	12.19*** (3.23)	9.42*** (3.41)	7.14** (3.15)	Yes	Yes
Scotland (UK)	13429	30.28*** (6.48)	30.28*** (5.27)	23.82*** (6.46)	Yes	Yes
Singapore	23208	23.83*** (3.21)	17.13*** (2.94)	19.75*** (3.26)	Yes	Yes

Notes: All estimates are based on multiply imputed test scores (plausible values); standard errors that account for clustering and for the sampling design are presented in parentheses and italics. *p<0.1; **p<0.05; ***p<0.01. Grade-and-age effects for each subject are estimated using separate regressions. They correspond to the coefficient on the interaction term between a dummy identifying cases tested during a late-testing window and a dummy identifying the months of birth of students who would have been (or were) older if tested during a late-testing window. See Equation 1 for details.

[1] The following covariates are included: girl, immigrant background, quarter of the index of socio-economic status (three dummies).

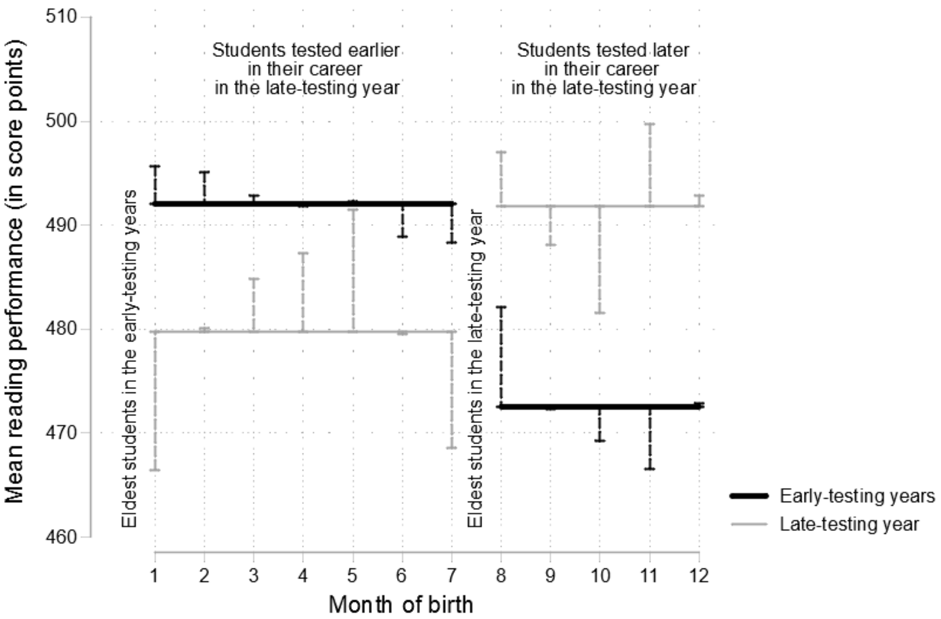
Source: PISA 2006, 2009, 2012, 2015 and 2018 datasets, `https://www.oecd.org/pisa/data/` (accessed on 17 May 2021).

tested later in their school career, perform at higher levels, while students who are tested earlier in their school career perform at lower levels. Indeed, this pattern is observed in Figure 3. The estimates in the top panel of Table 1 correspond to the sum of the vertical distances between the two sets of parallel horizontal lines.

Estimates that include controls for gender, socio-economic status, and immigrant background are presented below estimates without these controls; we focus on the estimates in the lower panel in our discussion. The differences between the two sets of estimates are relatively minor, and may have two causes. First, while the variables used in the first set of estimates are available for the full PISA samples, information on socio-economic status and immigrant background is based on self-report questionnaire and may be missing for a small fraction of students (we use list-wise deletion to deal with such missing values); a comparison of the number of observations however shows that this reduction is very minor (ranging from 2% in Austria and Singapore to about 6% in Brazil). Second, the socio-demographic characteristics of students may correlate with the difference-in-difference dummy in the presence, for example, of selective drop-out. The sensitivity of results to this potential selection bias is discussed in Section 5.2.

The estimates in Table 1 imply that students' test scores in PISA increase, over a full school year, on average by between 24 and 31 score points (depending on the subject) in Austria and Scotland, or about one-fourth of a standard deviation . The corresponding grade gain estimates lie between 17 and 24 score points in Singapore, or about one fifth of a standard deviation; between

Figure 3: Mean performance in reading in Austria, by month of birth and year



Notes: Horizontal lines indicate the average performance of students born in the corresponding months, by year. Drop lines show the mean performance estimate associated with each single month. In Austria, 2006, 2009, 2012 and 2018 were early-testing years, and 2015 was a late-testing year.
Source: PISA 2006, 2009, 2012, 2015 and 2018 datasets, `https://www.oecd.org/pisa/data/` (accessed on 17 May 2021).

11 and 15 score points, in Brazil; and between 7 and 12 score points, in Malaysia. The grade gain is less precisely estimated in Scotland, compared to the four remaining countries, due to the smaller sample size (Annex Table A.2 breaks down the sample size for each country by year; the precision of the difference-in-difference estimator mainly depends on the number of observations in the late-testing year).

4.1 Subgroup differences

In all five countries examined in this paper, at age 15 years, boys' performance lags behind girls' performance in reading, but boys score higher than girls in mathematics in Austria, Brazil and Scotland, while gender differences in science tend to be smaller than those observed in either mathematics or reading. In all three subjects, socio-economically disadvantaged students score below their more advantaged peers (see Annex Table A.3). To what extent are these gender differences and socio-economic gaps widening at age 15? Or do they rather reflect students' earlier learning experiences and/or different levels of skill at school start?

Table 2 reports the jointly estimated grade-and-age effects for each subgroup, as well as the difference between them (these correspond, respectively, to δ^0, $\delta^0 + \delta^1$ and δ^1 in Equation 2). Before commenting on the results, it must be noted that such a triple-difference estimator (δ^1) can be expected to have limited power in identifying differences in grade-and-age effects across subgroups, considering the magnitude of standard errors affecting the main analysis in Table 1. Only major differences in the grade gain across subgroups can be detected in the PISA samples used in this analysis.

The grade gain differs significantly across boys and girls in Scotland, with boys showing larger grade-and-age effects around the age of 15 compared to girls in all three subjects (Table 2). The corresponding gender differences in the remaining four countries are not significant and closer to 0.

Table 2: Grade-and-age effects by subgroup and between-group difference

Country	Maths			Reading			Science		
				Grade-and-age effects					
Gender									
	Boy	Girl	Diff.	Boy	Girl	Diff.	Boy	Girl	Diff.
Austria	25.44***	24.01***	−1.44	25.31***	33.52***	8.21	22.44***	26.82***	4.38
	(4.38)	(3.77)	(5.22)	(4.69)	(4.08)	(6.03)	(4.18)	(3.64)	(5.14)
Brazil	15.39***	8.46**	−6.93	19.01***	11.28**	−7.74	15.33***	8.73**	−6.60
	(4.94)	(4.25)	(6.21)	(4.74)	(5.12)	(6.84)	(4.65)	(3.99)	(6.07)
Malaysia	12.34***	12.01**	−0.33	9.17*	9.50*	0.33	4.98	9.15**	4.17
	(4.39)	(4.67)	(6.41)	(4.71)	(5.08)	(7.04)	(4.60)	(4.53)	(6.59)
Scotland (UK)	41.99***	19.26**	−22.73**	41.04***	20.46***	−20.59**	36.73***	11.50	−25.22***
	(7.44)	(8.83)	(9.97)	(7.86)	(6.09)	(9.25)	(8.58)	(7.40)	(9.39)
Singapore	23.48***	24.18***	0.70	15.39***	18.97***	3.57	19.07***	20.47***	1.39
	(4.62)	(4.38)	(6.34)	(4.37)	(4.55)	(6.77)	(4.67)	(4.38)	(6.37)
Socio-economic status (ESCS)									
	Low ESCS	High ESCS	Diff.	Low ESCS	High ESCS	Diff.	Low ESCS	High ESCS	Diff.
Austria	29.77***	19.73***	−10.03	35.21***	23.72***	−11.49*	28.87***	20.41***	−8.46
	(5.09)	(3.97)	(6.64)	(4.84)	(4.19)	(6.43)	(4.27)	(4.06)	(5.85)
Brazil	10.51***	12.87**	2.36	9.76*	19.99***	10.24	8.50**	15.00***	6.49
	(4.05)	(5.13)	(6.36)	(5.44)	(4.98)	(7.64)	(3.96)	(4.43)	(5.81)
Malaysia	9.27**	14.91***	5.64	7.99*	10.59**	2.60	4.29	9.82**	5.52
	(4.30)	(4.81)	(6.46)	(4.55)	(5.12)	(6.91)	(4.08)	(4.95)	(6.58)
Scotland (UK)	32.40***	28.81***	−3.59	37.47***	23.39***	−14.08*	28.80***	19.17**	−9.63
	(9.29)	(7.22)	(10.60)	(6.76)	(6.75)	(8.39)	(7.90)	(7.72)	(8.71)
Singapore	18.66***	28.83***	10.17	14.40***	19.77***	5.37	17.50***	21.85***	4.35
	(4.68)	(4.75)	(6.92)	(4.21)	(4.38)	(6.28)	(4.71)	(4.82)	(6.98)

Notes: All estimates are based on multiply imputed test scores (plausible values); standard errors that account for clustering and for the sampling design are presented in parentheses and italics. *p<0.1; **p<0.05; ***p<0.01. Each country and subject corresponds to a separate regression; the number of observations is the same across subjects, and is reported in Table 1 (bottom panel). Grade-and-age effects for subgroups correspond to coefficients on triple interaction terms between a dummy identifying cases tested during a late-testing window, a dummy identifying the months of birth of students who would have been (or were) older if tested during a late-testing window, and a dummy identifying the subgroup; the difference between the reported grade-and-age effects is also reported to allow testing for statistical significance. All regressions also include subgroup-specific year dummies and month-of-birth dummies (see Equation 2) as well as all control variables (girl, immigrant background, quarters of the index of socio-economic status).

[1] "Low ESCS" ("High ESCS") refers to students in the bottom (top) half of the country's distribution of the index of economic, social and cultural status.

Source: PISA 2006, 2009, 2012, 2015 and 2018 datasets, `https://www.oecd.org/pisa/data/` (accessed on 17 May 2021).

This suggests that in Scotland, boys reduce the gap in reading performance between the ages of 15 and 16, and widen the (small) gaps in mathematics performance. It is interesting to note that in most countries and economies, gender gaps in literacy among young adults observed in the Programme for the International Assessment of Adult Competencies (PIAAC) tend to be smaller than reading gaps observed in PISA among 15-year-olds, while numeracy gaps observed in PIAAC tend to be wider than the mathematics gap in PISA (Borgonovi, Choi and Paccagnella 2021). The fact that in Scotland the grade gain for boys around 15 years is somewhat larger than for girls in both reading and mathematics is consistent with this otherwise puzzling result; at the same time, the fact that such gender differences are not observed in the remaining countries suggests that the evolution of gender gaps is sensitive to institutional differences.

Finally, differences related to socio-economic status are, in general, non-significant (Table 2). The negative point estimates in Austria and Scotland, which are significant at the 10% level in reading, suggest that in these countries, the proficiency of disadvantaged children increases over one year of schooling (and age) at least as much as that of children from more advantaged families (who tend to be more proficient to start with). Schooling, in other words, does not reinforce pre-existing inequalities and may instead contribute to reducing socio-economic gaps. In contrast, in Brazil, Malaysia and Singapore, the grade-gain estimates are larger for more advantaged students than for their less advantaged peers, but not significantly so.

5 Assessing the strength of the common-trend assumption

The major assumption behind the identification strategy used in this paper is one of common (or parallel) trends across students born in different months.

It is not possible to formally test the common-trend assumption, but it is possible to corroborate it with further evidence. A first, indirect way of testing this assumption is to compare the trends for years in which there has been no change in testing dates, e.g. between 2006-2012 and 2018 in Austria, or between 2006 and 2015 in Scotland. If trends between these years are parallel, it is more likely that trends between the late-testing year and the early-testing years would also have been parallel in the absence of a change in the testing period.[7]

A second test to corroborate, more specifically, the absence of selection effects that could confound the age- and grade-differences associated with the difference-in-difference indicator in Equation 1, consists in comparing changes in the *observed* composition of the sample (in terms of gender, socio-economic status or immigrant background) across months of birth and across early- and late-testing years. If the groups defined by months of birth remain balanced, over the years, in terms of observable characteristics, this is more likely to be the case for unobservable characteristics as well.

A third test consists of comparing the differences in performance between early- and late-testing years across made-up month-of-birth groups: groups among which one would expect (based on Equation 1) such differences to be identical (i.e. to estimate a pseudo difference-in-difference). For example, in Austria and Scotland, one can focus on students born early in the calendar year only, and compare those born between January and March to those born between April and July. Both groups are expected to be affected equally by a change in testing dates, and any differences would therefore reflect some sort of violation of a common-trend assumption (in Brazil, Malaysia and Singapore, one can focus on students born later in the calendar year, which constitute the larger group, in order to maximise power). Finally, it is possible to test the extent to which results are driven by a single month of birth (and therefore, possibly, by month-of-birth-specific trends), rather than by a consistent pattern observed across all months that are similarly affected by the change in testing dates, by estimating Equation 1 using only the observations from 11 out of 12 months (leave-one-out estimator).

[7] This is similar to applications of difference-in-difference estimators that test the assumption of parallel trends (in the absence of an observed policy change) by showing "underlying" trends prior to the policy change (Angrist and Pischke 2010, 14-15).

5.1 Parallel trends

To corroborate the hypothesis of parallel trends (in the absence of changes in testing dates), two "placebo" differences-in-differences are reported.

The first restricts the sample to up to four assessment years in which the testing occurred on the same dates (i.e., excluding the late-testing year for each country). Under the assumption of common trends by month of birth, the interaction term between a PISA 2018 dummy (PISA 2015 for Scotland, where 2018 is the late-testing year) and being born towards the end of the calendar year (i.e., after the cut-off month highlighted in Figure 1; $m_i > M$ in Eq. 1) should not be significant in these regressions. Results, shown in Table 3 (Panel A), confirm that this is the case: only one the 15 estimated parameters appears to be significant, at the 10% level, which is in line with the proportion expected in standard test theory. The second placebo test restricts the estimation sample to students born in months such that they were all affected in the same direction by the change in testing period (The largest of the possible groups consists of students born in January through July in Austria and Scotland, in May through December in Brazil and Singapore, and in April through December in Malaysia). Among these students, it further distinguishes two made-up groups of approximately equal size (e.g. January-March vs. April-July in Austria and Scotland). Under the assumption of common trends by month of birth, the results of these groups should not diverge significantly in the late-testing year. Panel B in Table 3 shows, indeed, only few significant differences.[8]

5.2 Absence of selection effects

Enrolment rates in secondary education typically decrease at each grade level and as students get older. Because the grade levels attended by PISA students change when the testing dates change, it is important to investigate whether the composition of the student cohort also changes in ways that could confound the identification of grade gains.

For example, the grade-gain estimate for Austria corresponds mostly to the transition between Grades 9 and 10. By Grade 9, students in Austria have already started upper secondary education and are tracked into a general academic track or a number of vocational tracks. The most typical vocational tracks are school-based and begin in Grade 9, but students can also attend a pre-vocational year in Grade 9 before starting an apprenticeship by Grade 10. In this case, they only attend a part-time vocational school together with workplace-based vocational training. All types of schooling (part-time or full-time, general and vocational) are represented in the PISA sample, but schooling is compulsory only until age 15 in Austria (Salchegger and Suchań 2017). It is therefore possible that the composition of the student cohort changes around this age in the transition between grade levels and depending on whether the cohort is observed at the beginning or towards the end of a school year.

A test of the presence of selection effects is shown in Table 4. The balancing tests are performed with the same difference-in-difference estimator used to identify grade-and-age effects, where the dependent variable (test scores) has been replaced by one of the covariates (gender, immigrant background or quartile of socio-economic status).[9] The point estimates are close to zero and rarely statistically significant; however, at least for socio-economic status (the variable most closely associated with student performance), the small difference is such that students who are older, and more advanced in their schooling, tend to have slightly higher status in Austria and Scotland. While we interpret this difference as reflecting random sampling variation, the sign of this difference explains why grade gains that are estimated in a difference-in-difference regression with covariates

[8] In Scotland, both "placebo effects" (Panel A and Panel B) appear significant for science, but not for reading and mathematics. We consider this inconsistent pattern across subjects to be reassuring about the fact that age-and-grade effects reported in Tables 1 and 2 are not confounded by a systematic issue affecting our identification strategy for Scotland.

[9] This test assesses the presence of selection effects on observable sample characteristics; it cannot directly test the presence of residual selection effects on unobserved characteristics, after controlling for such observable characteristics

Table 3: Parallel trends by month of birth

Panel A. *Are performance changes different for those born towards the end of the calendar year, when the testing date remains the same?*

Country	Nb of obs.	Maths	Reading	Science	Year fixed effects	Month-of-birth dummies
Austria	23074	1.38 (3.92)	3.09 (3.72)	2.82 (3.87)	Yes	Yes
Brazil	73163	−4.26 (3.36)	−1.10 (2.74)	0.10 (3.02)	Yes	Yes
Malaysia	11308	2.51 (3.84)	−3.01 (4.27)	1.60 (4.02)	Yes	Yes
Scotland (UK)	11131	5.91 (4.21)	5.37 (4.26)	7.53^{*} (4.55)	Yes	Yes
Singapore	18337	−3.56 (3.60)	−2.70 (4.00)	−4.49 (3.72)	Yes	Yes

Panel B. *Are performance changes in the late-testing year different for students born in months which are equally affected by the change in the testing schedule?*

Country	Nb of obs.	Maths	Reading	Science	Year fixed effects	Month-of-birth dummies
Austria	17368	3.30 (3.75)	8.33^{**} (4.20)	4.95 (3.99)	Yes	Yes
Brazil	55227	5.97 (4.55)	7.27 (4.58)	2.80 (3.99)	Yes	Yes
Malaysia	12384	5.49^{*} (3.22)	5.65 (3.53)	4.08 (3.22)	Yes	Yes
Scotland (UK)	8065	3.59 (7.94)	4.73 (5.52)	13.52^{**} (6.41)	Yes	Yes
Singapore	16165	−0.36 (4.38)	−1.05 (4.28)	−1.58 (4.65)	Yes	Yes

Notes: All estimates are based on multiply imputed test scores (plausible values); standard errors that account for clustering and for the sampling design are presented in parentheses and italics. *p<0.1; **p<0.05; ***p<0.01. Placebo effects for each subject are estimated using separate regressions. In Panel A, they correspond to the coefficient on the interaction term between a dummy for PISA 2018 (PISA 2015 in Scotland) and a dummy identifying the months of birth of students who would have been older if tested during a late-testing window; the actual year in which PISA was conducted at a different date is excluded for each country from the sample, so that during all years included in the placebo regressions, students were actually tested in the same months. See Equation 1 for details. In Panel B, only students born in January through July (Austria and Scotland), in May through December (Brazil and Singapore) or in April through December (Malaysia) are included in the estimation sample; all students are therefore expected to be equally affected, in terms of age and length of schooling, by the change in testing period. Placebo effects correspond to the coefficient on the interaction term between a dummy for the late-testing year and a dummy identifying students born in April through July (Austria and Scotland), or in September through December (Brazil, Malaysia and Singapore). See Equation 1 for details.
Source: PISA 2006, 2009, 2012, 2015 and 2018 datasets, `https://www.oecd.org/pisa/data` (accessed on 17 May 2021).

tend to be slightly smaller than grade-gain estimates without controls for covariates in these two countries (Table 1).

Table 4: Absence of selection bias on grade-gain estimates.
Difference in sample covariates associated with one additional year of schooling and age

Country	Nb of obs.	Girl	Immigrant back-ground	High ESCS[1]	Year fixed effects	Month-of-birth dummies
Austria	29422	0.31 (1.57)	0.07 (1.52)	2.05 (1.62)	Yes	Yes
Brazil	77559	0.95 (2.04)	−0.09 (0.44)	−0.61 (1.68)	Yes	Yes
Malaysia	15921	0.29 (1.79)	0.03 (0.59)	−2.10 (2.16)	Yes	Yes
Scotland (UK)	13429	2.85* (1.67)	2.22** (0.91)	1.89 (2.25)	Yes	Yes
Singapore	23208	−0.23 (1.38)	−0.06 (1.12)	−0.94 (1.49)	Yes	Yes

Notes: Standard errors that account for clustering and for the sampling design are presented in parentheses and italics. *p<0.1; **p<0.05; ***p<0.01 Each reported coefficient is expressed as a percentage-point difference and is estimated using separate regressions. Selection effects correspond to the coefficient on the interaction term between a dummy identifying cases tested during a late-testing window and a dummy identifying the months of birth of students who would have been (or were) older if tested during a late-testing window. See Equation 1 for details.
[1] High ESCS refers to students in the top half of the country's distribution of the index of economic, social and cultural status.
Source: PISA 2006, 2009, 2012, 2015 and 2018 datasets, `https://www.oecd.org/pisa/data/` (accessed on 17 May 2021).

5.3 Leave-one-out analysis

The final robustness check consists of estimating Equation 1 on 12 different subsamples, each defined by excluding students born in a particular month from the main sample. If results reported in Table 1 are driven by a change affecting only a particular month of birth, one would expect these alternative difference-in-difference estimates to show wide variation. In contrast, if the results are driven by the age- and length-of-schooling variation that is common to several months, results should not vary much across the 12 estimates. This is what Table 5 shows.

6 Discussion

The present article quantifies the learning gain that results from an additional year of schooling in secondary schools, using data from a well-known large-scale international assessment. Its original identification strategy overcomes the limitations of previous studies that relied on a regression-discontinuity design and provides first-of-its-kind comparative evidence on the effectiveness of schooling around the age of 15 years.

The estimates reported in the present article indicate that the typical grade gain for 15-year-old students varies widely across countries (Figure 4). Among the five countries considered in this study, the smallest grade gains are observed in Malaysia: on subject-specific standardised scales, where 100 score points correspond to one standard deviation in an international reference population of 15-year-old students, the yearly gains in Malaysia are of only 7 score points in science, 9 score points in reading and 12 score points in mathematics. Brazil's estimates (12, 15 and 12 score points, respectively) are higher than Malaysia's in reading and science, but not significantly

Table 5: Robustness of grade-gain estimates

Country	Grade-and-age-effect (min-max)					
	Maths		Reading		Science	
	min	max	min	max	min	max
Austria	23.60	29.64	28.12	33.32	23.74	27.50
Brazil	9.07	12.81	12.43	16.61	9.86	13.13
Malaysia	10.09	12.20	6.30	11.02	3.87	8.89
Scotland (UK)	30.90	34.56	31.48	36.39	24.97	28.80
Singapore	20.95	25.45	13.89	18.28	16.92	19.87

Notes: The table reports the range (minimum - maximum) of estimates across 12 samples, each defined by excluding one month of birth from the main estimation sample. All estimates include month-of-birth dummies and year-fixed effects. Grade-and-age effects for each subject are estimated using separate regressions. They correspond to the coefficient on the interaction term between a dummy identifying cases tested during a late-testing window and a dummy identifying the months of birth of students who would have been (or were) older if tested during a late-testing window. See Equation efeq:mainequa for details.

Source: PISA 2006, 2009, 2012, 2015 and 2018 datasets, `https://www.oecd.org/pisa/data` (accessed on 17 May 2021).

so, when considering the statistical uncertainty associated with each estimate.[10] Compared to 15-year-olds in Brazil and Malaysia, students of the same age in Austria, Scotland (United Kingdom) and Singapore appear to make significantly stronger progress per year. The yearly learning gain around the age of 15 in mathematics (25, 30 and 24 score points, respectively, in Austria, Scotland and Singapore) and science (25, 24 and 20 score points) is similar across all three countries: the small differences are not statistically significant at the 5% level. In reading, in contrast, students in Austria and Scotland appear to make stronger gains (29 and 30 score points, respectively) compared to students in Singapore (17 score points), whose annual reading gain is not significantly larger than that of students in Brazil and Malaysia.

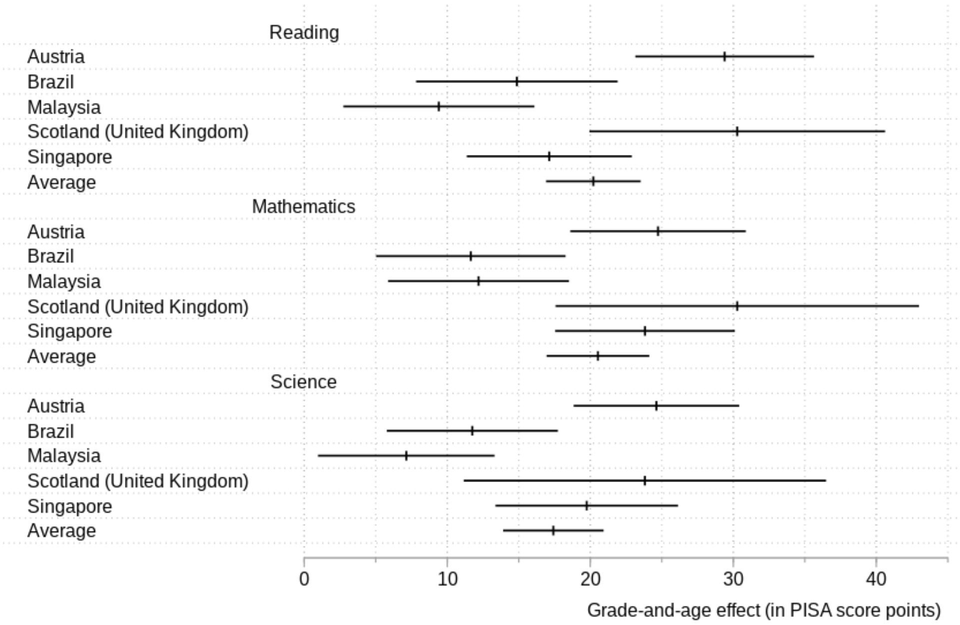
When considered in the context of the average performance of students in PISA, these results imply that students in Austria, Scotland and Singapore not only score (much) higher than students in Brazil and Malaysia, but also that international learning gaps continue to widen at age 15. As a result, any test taken in school may under-estimate the skills gap between the adult populations of high- and middle-income countries. There is only limited evidence of students catching up at age 15: students in Austria and Scotland appear to be narrowing the gap to students in Singapore in reading, in particular.

Somewhat more positive news emerges when considering the learning dynamics within countries; in general, no significant sub-group differences were found, including by students' socio-economic status: meaning that the wide socio-economic gaps in test scores do reflect, to a significant extent, pre-existing inequalities, including differences in cultural norms transmitted by parents (De Philippis and Rossi 2020), rather than the effect of inequitable learning opportunities in secondary education. This relative stability of socio-economic gaps is observed across a variety of systems, including systems with early tracking by ability (Austria, Singapore), and systems where the most advantaged and the most disadvantaged students are highly segregated (Brazil, Malaysia).

The average grade effect for 15-year-olds reported in the present study can be used as a benchmark for assessing the practical significance of other performance differences observed in PISA. For example, in 2018, the difference in mean scores in mathematics between the United States (478 points) and the United Kingdom (502 points) was about the size of the typical test-score

[10]The significance of cross-country differences can be assessed by taking advantage of the independence of national samples in PISA. The standard error for the difference between two estimators is simply the square root of the sum of the variances of the two estimators.

Figure 4: Grade-and-age effects across 5 countries



Notes: The following differences are statistically significant at the 5% level (two-sided p-values under the null of no difference are in parentheses):

Reading Austria - Brazil (.002), Austria - Malaysia (.000), Austria - Singapore (.005), Brazil - Scotland (.016), Malaysia - Scotland (.001), Scotland - Singapore (.030).

Mathematics Austria - Brazil (.004), Austria - Malaysia (.005), Brazil - Scotland (.011), Brazil - Singapore (.009), Malaysia - Scotland (.012), Malaysia - Singapore (.010)

Science Austria - Brazil (.002), Austria - Malaysia (.000), Malaysia - Scotland (.020), Malaysia - Singapore (.005).

Source: Table 1, Panel B.

gap observed in high-income countries between students who are one grade level apart, around the age of 15 (OECD 2019a); as was the gender gap in reading (30 score points, on average across OECD countries) (OECD 2019b). But it would take students in the bottom 25% of socioeconomic status, who in reading score on average 89 points lower than students in the top 25%, several years of schooling to reach the current level of their more advantaged peers.

Over the coming years, several international assessments will be conducted and their results will be closely scrutinised to understand the impact of the disruptions to regular schooling induced by the COVID pandemic. In this respect, the five countries analysed in the present study present an interesting variety. Brazil was one of the hardest-hit countries in the first year of COVID; according to UNESCO data, between 1 March 2020 and 30 June 2021, schools in Brazil were fully closed for 38 weeks and partially closed (meaning that they remained closed for a sizeable proportion of students) for further 19 weeks. Outside of academic breaks, regular schooling was in place only during a few weeks at the beginning of the period. In Malaysia, schools were closed for 28 weeks and partially closed for further 12 weeks. Meanwhile, in Singapore, school closed only for 4 weeks due to COVID, and were partially closed for 9 weeks. Austria and the United Kingdom are in an intermediate position: schools remained closed for about 15 weeks, with partial closures during 24 and 11 weeks respectively (data are not available separately for Scotland) (UNESCO 2021). In order to interpret the difference with pre-COVID assessments, the extent to which learning was disrupted – of which school closures are an indicator – must be considered; but it is equally important to consider differences in school productivity prior to the pandemic, highlighted in the present article. In countries where productivity was low to start with, major disruptions may have had less severe consequences than milder disruptions in high-productivity countries.

An aspect on which future data may also shed new light is the relative importance of school instruction and of other life experiences for skill acquisition. In the present article, this question could not be addressed, and all estimates reflect the combined influence of schooling, maturity, and other sources of cognitive development in the life of 15-year-olds on the skills assessed in PISA tests. Previously, this question has been examined in the literature based on seasonal patterns in test scores,[11] by examining the long-term consequences of reforms that varied the number of school days per year, without changing the number of years of schooling (Fischer, et al. 2019), or by exploiting the random variation in the date of cognitive tests taken in preparation for military service (Carlsson, et al. 2015). All kinds of studies provide some evidence in support of the importance of school education and, indirectly, for the interpretation of grade-gain differences (such as those shown in Figure 4) as reflecting underlying differences in the productivity of schooling. It is clear that the implications of school closures for learning and, beyond, for the skills and human capital of the affected cohorts, will vary depending on the importance of (in-person) schooling in the technology of skill acquisition.

[11] Several studies in the United States, summarised in an influential meta-analysis, have highlighted a "summer learning loss", i.e. an average fall in test scores during the summer break in elementary school (Cooper, et al. 1996). This suggests that there are no age/maturity effects on test scores or that these might even be negative. However, more recent studies have suggested that this finding may suffer from methodological flaws. Indeed, when more comparable tests and better scaling techniques are used to examine seasonal patterns of learning, the finding of a "learning loss" during the early school years does not always replicate (von Hippel and Hamrock 2019). A more recent study, using a large dataset spanning eight grades of schooling (Grades 1 to 8), has found that test scores decline during the summer months, but that this average loss decreases as students move from elementary to lower secondary grades (Atteberry and McEachin 2020).

References

Anders, Jake, John Jerrim, and Andrew McCulloch. 2016. "How Much Progress Do Children in Shanghai Make Over One Academic Year? Evidence From PISA." *AERA Open* (SAGE Publications) 2 (4): 233285841667884. doi:10.1177/2332858416678841.

Angrist, Joshua D, and Jörn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics." *Journal of Economic Perspectives* (American Economic Association) 24 (2): 3-30. doi:10.1257/jep.24.2.3.

Atteberry, Allison, and Andrew McEachin. 2020. "School's Out: The Role of Summers in Understanding Achievement Disparities." *American Educational Research Journal* (American Educational Research Association (AERA)) 000283122093728. doi:10.3102/0002831220937285.

Avvisati, Francesco, and François Keslair. 2014. *REPEST: Stata module to run estimations with weighted replicate samples and plausible values.* Statistical Software Components S457918, Boston College Department of Economics. Accessed 01 05, 2021.

Avvisati, Francesco, and Pauline Givord. 2021. "How much do 15-year-olds learn over one year of schooling? An international comparison based on PISA" , *OECD Education Working Papers*, No. 257. OECD Publishing, Paris. doi:10.1787/a28ed097-en.

Bedard, K., and E. Dhuey. 2006. "The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects." *The Quarterly Journal of Economics* (Oxford University Press (OUP)) 121 (4): 1437-1472. doi:10.1093/qje/121.4.1437.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2011. "Too Young to Leave the Nest? The Effects of School Starting Age." *Review of Economics and Statistics* (MIT Press - Journals) 93 (2): 455-467. doi:10.1162/rest_a_00081.

Bond, Timothy N., and Kevin Lang. 2013. "The Evolution of the black–white test score gap in grades K-3: The fragility of results" *Review of Economics and Statistics* 95 (5): 1468–1479. doi:10.1162/REST_a_00370.

Borgonovi, Francesca, Alvaro Choi, and Marco Paccagnella. 2021. "The evolution of gender gaps in numeracy and literacy between childhood and young adulthood." *Economics of Education Review* (Elsevier BV) 82: 102119. doi:10.1016/j.econedurev.2021.102119.

Braun, Henry, and Matthias von Davier. 2017. "The use of test scores from large-scale assessment surveys: psychometric and statistical considerations" *Large-scale Assessments in Education* 5(17). doi: 10.1186/s40536-017-0050-x.

Carlsson, Magnus, Gordon B. Dahl, Björn Öckert and Dan-Olof Rooth. 2015. "The Effect of Schooling on Cognitive Skills". *The Review of Economics and Statistics.* 97 (3): 533–547. doi:10.1162/REST_a_00501

Cooper, Harris, Barbara Nye, Kelly Charlton, James Lindsay, and Scott Greathouse. 1996. "The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic Review." *Review of Educational Research* (American Educational Research Association (AERA)) 66 (3): 227-268. doi:10.3102/00346543066003227.

Crawford, Claire, Lorraine Dearden, and Ellen Greaves. 2014. "The drivers of month-of-birth differences in children's cognitive and non-cognitive skills." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* (Wiley) 177 (4): 829-860. doi:10.1111/rssa.12071.

De Philippis, Marta, and Federico Rossi. 2020. "Parents, Schools and Human Capital Differences Across Countries." *Journal of the European Economic Association* (Oxford University Press (OUP)) 19 (2): 1364-1406. doi:10.1093/jeea/jvaa036.

Dearden, Lorraine, Claire Crawford, and Costas Meghir. 2010. "When you are born matters: the impact of date of birth on educational outcomes in England." In *Working Paper Series.* Institute for Fiscal Studies. doi:10.1920/wp.ifs.2010.1006.

Fischer, Martin, Martin Karlsson, Therese Nilsson, and Nina Schwarz. 2019. "The Long-Term Effects of Long Terms – Compulsory Schooling Reforms in Sweden." *Journal of the European Economic Association* (Oxford University Press (OUP)) 18 (6): 2776-2823. doi:10.1093/jeea/jvz071.

Givord, Pauline. 2020. "How a student's month of birth is linked to performance at school: New evidence from PISA." *OECD Education Working Papers*, No.221. OECD Publishing, Paris. doi:10.1787/822ea6ce-en.

Gomes, Matheus, Guilherme Hirata, and Joao Batista Araujo e Oliveira. 2020. "Student composition in the PISA assessments: Evidence from Brazil." *International Journal of Educational Development* (Elsevier BV) 79: 102299. doi:10.1016/j.ijedudev.2020.102299.

Hanushek, Eric A., Mark Piopiunik, and Simon Wiederhold (2019), "The value of smarter teachers international evidence on teacher cognitive skills and student performance", *Journal of Human Resources*, 54(4), 857-899.

Jacob, Brian, and Jesse Rothstein. 2016. "The measurement of student ability in modern assessment systems." *Journal of Economic Perspectives* 30(3) (2016): 85-108. doi:10.1257/jep.30.3.85.

Jones, S., Y. Schipper, S. Ruto, and R. Rajani. 2014. "Can Your Child Read and Count? Measuring Learning Outcomes in East Africa." *Journal of African Economies* (Oxford University Press (OUP)) 23 (5): 643-672. doi:10.1093/jae/eju009.

Kuzmina, Julia, and Martin Carnoy. 2016. "The effectiveness of vocational versus general secondary education." *International Journal of Manpower* (Emerald) 37 (1): 2-24. doi:10.1108/ijm-01-2015-0022.

Lavy, Victor (2015), "Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries", *Economic Journal*, 125, F397-F424. https://doi.org/10.1111/ecoj.12233

Luyten, Hans, and Bernard Veldkamp. 2011. "Assessing Effects of Schooling With Cross-Sectional Data: Between-Grades Differences Addressed as a Selection-Bias Problem." *Journal of Research on Educational Effectiveness* (Informa UK Limited) 4 (3): 264-288. doi:10.1080/19345747.2010.519825.

Luyten, Hans, Jules Peschar, and Robert Coe. 2008. "Effects of Schooling on Reading Performance, Reading Engagement, and Reading Activities of 15-Year-Olds in England." *American Educational Research Journal* (American Educational Research Association (AERA)) 45 (2): 319-342. doi:10.3102/0002831207313345.

Nagy, Gabriel, Oliver Lüdtke, Olaf Köller, and Jörg-Henrik Heine. 2017. "IRT-Skalierung der Tests im PISA-Längsschnitt 2012/2013: Auswirkungen von Testkontexteffekten auf die Zuwachsschätzung." *Zeitschrift für Erziehungswissenschaft* (Springer Science and Business Media LLC) 20 (S2): 229-258. doi:10.1007/s11618-017-0749-z.

OECD. 2020. "Characteristics of education systems." Chap. 34 in *Education at a Glance 2020: OECD Indicators.* OECD Publishing, Paris. doi:https://dx.doi.org/10.1787/4bb100a8-en.

-. 2017. *PISA 2015 Technical Report.* Accessed 07 31, 2017. http://www.oecd.org/pisa/data/2015-technical-report/.

-. 2019a. *PISA 2018 Results (Volume I): What Students Know and Can Do.* Paris: OECD Publishing.

-. 2019b. "PISA 2018 Results (Volume II): Where All Students Can Succeed." In *PISA*. OECD Publishing, Paris. doi:https://dx.doi.org/10.1787/b5fd1b8f-en.

Prenzel, Manfred, Jürgen Baumert, Werner Blum, Rainer Lehmann, Detlev Leutner, Michael Neubrand, Reinhard Pekrun, Jürgen Rost, and Ulrich Schiefele, . 2006. *PISA 2003: Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres.* Münster: Waxmann Verlag GmbH.

Rosa, Leonardo, Marcelo Martins, and Martin Carnoy. 2019. "Achievement gains from reconfiguring early schooling: The case of Brazil's primary education reform." *Economics of Education Review* (Elsevier BV) 68: 1-12. doi:10.1016/j.econedurev.2018.10.010.

Salchegger, Silvia, and Birgit Suchañ. 2017. "Was bedeutet es für den Geschlechterunterschied in der Mathematikkompetenz bei PISA, wenn dem Schulsystem leistungsschwache Jungen verloren gehen?" *Zeitschrift für Bildungsforschung* (Springer Science and Business Media LLC) 8 (1): 81-99. doi:10.1007/s35834-017-0190-7.

Schroeder, Carsten, and Shlomo Yitzhaki. 2020. "Exploring the robustness of country rankings by educational attainment" *Journal of Economics* 129: 271–296. doi:10.1007/s00712-019-00682-8.

Singh, Abhijeet. 2019. "Learning More with Every Year: School Year Productivity and International Learning Divergence." *Journal of the European Economic Association* (Oxford University Press (OUP)) 18 (4): 1770-1813. doi:10.1093/jeea/jvz033.

Tiumeneva, Yu. A., and Ju. V. Kuzmina. 2015. "The Difference That One Year of Schooling Makes for Russian Schoolchildren." *Russian Education & Society* (Informa UK Limited) 57 (4): 214-253. doi:10.1080/10609393.2015.1068567.

UNESCO. 2021. *Global dataset on the duration of school closures.* Vers. Last update: 30 June 2021. Accessed August 23, 2021. https://en.unesco.org/sites/default/files/duration_school_closures.csv.
von Hippel, Paul T., and Caitlin Hamrock. 2019. "Do test score gaps grow before, during, or between the school years? Measurement artifacts and what we can know in spite of them." *Sociological Science* (Society for Sociological Science) 6: 43-80. doi:10.15195/V6.A3.
Woessmann, Ludger. 2016. "The Importance of School Systems: Evidence from International Differences in Student Achievement." *Journal of Economic Perspectives* (American Economic Association) 30 (3): 3-32. doi:10.1257/jep.30.3.3.

A Appendix: Descriptive statistics

Table A.1: Descriptive statistics on PISA samples used

		Average		2006		2009		2012		2015		2018	
	Variable	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
Austria	Age	15.8	0.3	15.8	0.3	15.8	0.3	15.8	0.3	15.8	0.3	15.8	0.3
	Grade[1]	9.0	0.6	9.0	0.6	9.0	0.6	9.1	0.6	9.0	0.6	9.0	0.6
	Girl	49.8		49.1		51.0		50.1		49.5		49.2	
	High ISCED	50.8		50.9		47.7		48.0		52.7		54.8	
	Immigrant	17.6		13.2		15.2		16.5		20.3		22.7	
	Mathematics	500.4	95.1	505.5	98.1	495.9	96.1	505.5	92.5	496.5	94.8	498.5	93.9
	Reading	483.8	100.1	490.2	108.2	470.3	100.1	489.6	91.8	484.7	101.0	484.2	99.3
	Science	499.0	96.9	510.8	97.9	494.3	101.8	505.8	92.2	495.0	97.1	489.2	95.4
Brazil	Age	15.9	0.3	15.8	0.3	15.9	0.3	15.9	0.3	15.9	0.3	15.9	0.3
	Grade[1]	8.9	1.0	8.2	0.9	8.3	0.9	9.4	1.0	9.3	1.1	9.3	1.1
	Girl	52.1		53.8		53.1		51.9		51.5		50.0	
	High ISCED	33.4		35.9		30.6		25.3		29.4		45.6	
	Immigrant	1.1		2.4		0.8		0.7		0.8		0.6	
	Mathematics	380.9	85.6	369.5	92.0	385.8	81.2	388.5	78.2	377.4	89.0	383.3	87.8
	Reading	406.3	96.6	392.9	102.5	411.8	94.0	406.5	86.4	407.2	100.2	412.9	99.7
	Science	400.4	86.4	390.3	89.3	405.4	84.0	401.6	79.4	400.7	89.1	403.7	90.2
Malaysia	Age	15.8	0.3			15.8	0.3	15.8	0.3			15.8	0.3
	Grade[1]	9.2	0.3			9.2	0.5	9.2	0.2			9.2	0.2
	Girl	51.3				50.9		51.6				51.3	
	High ISCED	30.4				27.8		28.8				34.5	
	Immigrant	1.6				1.3		1.7				1.6	
	Mathematics	421.7	79.1			404.3	73.3	420.5	81.1			440.4	82.9
	Reading	409.0	83.0			413.8	80.5	398.2	83.7			414.9	84.8
	Science	426.5	77.1			422.2	75.8	419.5	78.6			437.7	76.8
Scotland (UK)	Age	15.7	0.3	15.7	0.3	15.7	0.3	15.7	0.3	15.7	0.3	15.8	0.3
	Grade[1]	10.6	0.4	10.6	0.3	10.6	0.3	10.6	0.3	10.5	0.3	10.6	0.5
	Girl	49.7		49.5		49.8		49.6		49.1		50.6	
	High ISCED	59.6		50.9		58.0		59.6		63.9		65.7	
	Immigrant	5.8		2.6		4.0		8.4		5.7		8.4	
	Mathematics	496.6	88.1	505.7	84.5	499.0	92.5	498.4	86.4	490.9	83.6	488.7	93.5
	Reading	500.5	92.4	498.8	95.6	500.1	94.2	506.1	86.7	493.3	90.5	504.0	95.2
	Science	505.9	95.3	514.7	99.9	514.2	95.7	513.4	89.4	497.0	94.7	490.4	96.9
Singapore	Age	15.8	0.3			15.7	0.3	15.8	0.3	15.8	0.3	15.8	0.3
	Grade[1]	9.1	0.4			9.1	0.6	9.1	0.4	9.1	0.4	9.2	0.3
	Girl	48.9				49.2		49.0		48.3		49.0	
	High ISCED	53.0				43.2		47.8		55.7		65.4	
	Immigrant	19.6				14.4		18.3		20.9		24.8	
	Mathematics	567.2	99.7			562.0	104.4	573.5	105.4	564.5	95.4	568.8	93.8
	Reading	538.2	101.6			525.9	97.5	542.2	100.9	535.0	99.1	549.5	108.9
	Science	549.9	102.3			541.7	104.0	551.5	104.2	555.5	103.8	551.1	97.2

Notes: Means and standard deviations (s.d.) of mathematics, reading and science scores are based on multiply imputed test scores (plausible values). "High ISCED" identifies students who reported that at least one parent completed a tertiary-level degree (ISCED 5A, 5B or 6). Means of binary variables ("Girl", "High ISCED" and "Immigrant") are reported in percentage points.

1. The number of completed grade levels is computed as the current grade, minus 1, plus the difference between the age of the student and his or her age at the beginning of the school year.

Source: PISA 2006, 2009, 2012, 2015 and 2018 datasets, `https://www.oecd.org/pisa/data/` (accessed on 17 May 2021).

Table A.2: Sample size by country and year

	Total	2006	2009	2012	2015	2018
Austria	30080	4927	6590	4755	7006	6802
Brazil	82458	9295	20127	19204	23141	10691
Malaysia	16307		4999	5197		6111
Scotland (UK)	14129	2444	2631	2945	3111	2998
Singapore	23620		5283	5546	6115	6676

Source: PISA 2006, 2009, 2012, 2015 and 2018 datasets, `https://www.oecd.org/pisa/data/` (accessed on 17 May 2021).

Table A.3: Mean performance in PISA, by gender and socio-economic status

		Mathematics		Reading		Science	
Country	Gender	mean	s.e.	mean	s.e.	mean	s.e.
Austria	boys	510.7	1.7	466.6	1.8	503.4	1.8
	girls	489.9	1.7	501.1	1.8	494.6	1.8
	Difference (g-b)	-20.8	2.2	34.5	2.4	-8.9	2.3
Brazil	boys	388.7	1.2	391.5	1.4	402.0	1.2
	girls	373.7	1.2	419.8	1.3	398.9	1.1
	Difference (g-b)	-15.0	1.0	28.2	1.0	-3.0	0.8
Malaysia	boys	418.7	2.0	391.7	1.9	421.9	1.9
	girls	424.5	1.9	425.4	1.8	430.8	1.8
	Difference (g-b)	5.8	2.0	33.6	1.6	8.9	1.8
Scotland (UK)	boys	503.1	1.9	489.1	1.8	508.8	1.9
	girls	489.9	1.7	511.9	1.5	503.0	1.7
	Difference (g-b)	-13.2	1.9	22.8	1.8	-5.8	2.1
Singapore	boys	568.0	0.9	525.2	0.9	550.9	0.9
	girls	566.4	1.1	551.7	1.0	548.9	0.9
	Difference (g-b)	-1.7	1.2	26.6	1.2	-2.0	1.2
	Quarters of	Mathematics		Reading		Science	
Country	ESCS	mean	s.e.	mean	s.e.	mean	s.e.
Austria	q1 (bottom)	454.7	2.1	436.9	2.2	450.2	2.1
	q2	488.6	1.8	471.1	1.8	487.1	1.8
	q3	512.7	1.7	497.2	1.6	512.0	1.6
	q4 (top)	547.9	1.7	534.0	1.7	550.0	1.6
	Difference (t-b)	93.2	2.6	97.0	2.7	99.7	2.6
Brazil	q1 (bottom)	344.9	1.4	370.0	1.6	366.0	1.2
	q2	366.1	1.1	392.8	1.5	385.6	1.1
	q3	383.9	1.3	411.0	1.5	403.7	1.4
	q4 (top)	431.7	2.1	455.6	2.0	450.0	1.9
	Difference (t-b)	86.8	2.6	85.6	2.6	84.0	2.2
Malaysia	q1 (bottom)	388.3	1.6	380.3	1.9	398.1	1.7
	q2	409.3	1.6	398.1	1.9	415.7	1.8
	q3	425.1	2.0	409.7	2.1	427.6	2.0
	q4 (top)	465.3	2.9	448.8	2.7	465.4	2.5
	Difference (t-b)	77.0	3.2	68.5	3.2	67.2	3.0
Scotland (UK)	q1 (bottom)	460.3	2.2	464.9	1.9	466.0	2.1
	q2	485.4	2.1	489.3	1.8	492.6	2.0
	q3	506.7	2.0	510.9	1.9	516.9	1.9
	q4 (top)	542.8	2.4	546.2	2.1	557.2	2.2
	Difference (t-b)	82.5	3.6	81.3	2.6	91.2	3.1
Singapore	q1 (bottom)	516.7	1.5	486.6	1.4	496.1	1.5
	q2	554.4	1.4	524.7	1.4	535.0	1.5
	q3	584.2	1.4	553.9	1.3	568.3	1.3
	q4 (top)	614.9	1.8	589.3	1.6	601.9	1.5
	Difference (t-b)	98.2	2.0	102.7	2.1	105.8	2.0

Notes: Means of Mathematics, Reading and Science scores are based on multiply imputed test scores (plausible values); standard errors (s.e.) account for clustering and for the sampling design. For each country, the average across all available years is presented. ESCS refers to the index of economic, social, and cultural status.
Source: PISA 2006, 2009, 2012, 2015 and 2018 datasets, `https://www.oecd.org/pisa/data/` (accessed on 17 May 2021).

Le LIEPP (Laboratoire interdisciplinaire d'évaluation des politiques publiques) est un laboratoire d'excellence (Labex) distingué par le jury scientifique international désigné par l'Agence nationale de la recherche (ANR). Il est financé dans le cadre des investissements d'avenir *(ANR-11-LABX-0091, ANR-11-IDEX-0005-02)* ***et de l'IdEx Université de Paris*** *(ANR-18-IDEX-0001).*

www.sciencespo.fr/liepp

A propos de la publication

Procédure de soumission :
Rédigé par un ou plusieurs chercheurs sur un projet en cours, le *Working paper* vise à susciter la discussion scientifique et à faire progresser la connaissance sur le sujet étudié. Il est destiné à être publié dans des revues à comité de lecture (peer review) et à ce titre répond aux exigences académiques. Les textes proposés peuvent être en français ou en anglais. En début de texte doivent figurer : les auteurs et leur affiliation institutionnelle, un résumé et des mots clefs.

Le manuscrit sera adressé à : liepp@sciencespo.fr

Directrice de publication :
Anne Revillard

Comité de redaction :
Sofia Cerda Aparicio, Andreana Khristova

Sciences Po - LIEPP
27 rue Saint Guillaume
75007 Paris - France
+33(0)1.45.49.83.61
liepp@sciencespo.fr