Organisation for Economic Co-operation and Development

EDU/WKP(2021)6

Unclassified **English text only**

29 June 2021

DIRECTORATE FOR EDUCATION AND SKILLS

The learning gain over one school year among 15-year-olds: An analysis of PISA data for Austria and Scotland (United Kingdom)

OECD Education Working Paper No. 249

Francesco AVVISATI (OECD) and Pauline GIVORD (INSEE-Liepp)

This working paper has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.

Francesco AVVISATI, Francesco.AVVISATI@oecd.org

JT03478937

OECD EDUCATION WORKING PAPERS SERIES

Comment on the series is welcome, and should be sent to edu.contact@oecd.org.

This working paper has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.

---

www.oecd.org/edu/workingpapers

Acknowledgements

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 from July to December 2020. We are also indebted to those who participated in the Applied Economics Work-in-Progress Seminar at the OECD (October 2020). Previous versions of this manuscript benefited from valuable feedback from Andreas Schleicher, Yuri Belfali, Tiago Fragoso, Miyako Ikeda and Marco Paccagnella, as well as from experts at the Federal Institute for Quality Assurance in the Austrian School System (IQS).

Abstract

This paper compares the learning gain over one year of schooling among 15-year-old students in Austria and Scotland (United Kingdom). 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, multiple cycles of PISA data are combined. The fact that Austria and Scotland moved their testing period across cycles is also exploited. The results are used to establish a benchmark for other performance differences observed in PISA, such as gender gaps, socio-economic gaps or between-country differences.

Table of contents

Tables

Figures

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. They are like a snapshot of student learning; to compare learning gains based on international assessments, one would need to build a time-lapse animation from a single snapshot.

This paper tries to address this challenge and provides rigorous causal estimates of the effect of an additional year of schooling and year of age on performance in two jurisdictions, based on common metrics for reading, mathematics and science learning established by the Programme for International Student Assessment (PISA). Such estimates constitute evidence of the relative effectiveness of learning systems around the age of 15 years. To the extent that the estimates validly represent what would have happened in 2020/21 in the absence of school closures, they also provide an upper-bound estimate for the "learning loss" that can be attributed to school closures during the coronavirus (COVID-19) crisis. Estimates of the grade gain based on PISA can also be used as a benchmark for other differences in performance, such as gender gaps, socio-economic gaps or between-country differences. They can also be used as a benchmark for the differences that will eventually be observed between the pre-COVID-19 assessments and the post-COVID-19 assessments, which will provide direct evidence of the extent to which education systems were able to maintain their "normal" productivity during those years. With due caution, such measures can also be used to express the potential gains from educational interventions and reforms in approximate years-of-schooling equivalents.

This study shows that in both Austria and Scotland (United Kingdom), students' yearly learning progress around the age of 15 is equivalent to about one-fourth of a standard deviation in students' test scores. This study also suggests that the learning gains largely reflect the effect of attending school for one additional year (rather than age or maturity effects) and that, in Austria – where student tracking begins before age 15 – the gains in reading, mathematics and science literacy over one year of schooling are markedly larger in more academically-oriented tracks compared to vocational tracks. In general, however, socio-economic disparities in learning do not widen around the age of 15.

All estimates in this paper are interpreted in terms of a simple education production function in which students' learning is a function of their age, their length of schooling and their age at school entry, along with other factors. Because at any point in time, these three variables are bound by a simple additive relationship, it is usually not possible to identify their distinct contribution in a cross-sectional design. Furthermore, the observed length of schooling and the actual age at school entry are possibly endogenous, influenced as they are by prior performance and other (mostly unobserved) determinants of student learning. To interpret the observed associations between length of schooling, school-entry age and learning as reflecting causality, one must rely only on exogenous sources of variation in these variables.

We use the change, across different cycles, in the time of the year when PISA was conducted in these two jurisdictions to identify the joint effect of age and length of schooling on learning on representative populations of students. In PISA, the target population is defined by a 12-month range of age, rather than by a grade level (as is usual in most national assessments and as is the case in other international large-scale

assessments such as the Trends in International Mathematics and Science Study [TIMSS]). In any given cycle, the birthdates of PISA students span all 12 months; however, when the testing date changes to an earlier time in the school year (as was the case in Austria, in 2015, and in Scotland, in 2018), the month of birth of the eldest eligible students also changes. 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. 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.

After identifying the average learning gain, differences in the joint effect of schooling and age across subgroups of students, which differ in their schooling experience at the age of 15, are explored.

This paper is organised as follows:

- The remainder of this section provides a selective overview of prior studies on the grade gain and related literature;
- Section 2 introduces the characteristics of PISA data and samples used in this paper;
- Section 3 describes the identification strategy;
- Section 4 presents the difference-in-difference results and robustness checks;
- Section 5 presents additional results from subgroup analyses; and
- Section 6 discusses the results.

1.1. Related literature

The present paper is related to three main strands in the multi-disciplinary literature on education and learning.

A first strand comprises the small literature comparing the productivity of a grade of schooling across countries (Singh, 2019[1]; Jones et al., 2014[2]), and the more extensive literature that quantifies the grade gain based on longitudinally-linked assessments, within a single education system (Prenzel et al., 2006[3]; Nagy et al., 2017[4]; Andrabi et al., 2011[5]; Chetty, Friedman and Rockoff, 2014[6]; Kane and Staiger, 2008[7]). Among the former comparative studies, several contributions have previously used cross-sectional data from PISA (as in the present study) or from other international large-scale assessments to identify the contribution of schooling to the grade gain. Luyten, Peschar and Coe (2008[8]) compare, using PISA 2000 data from the United Kingdom (excluding Scotland), students who are born just before and immediately after the cut-off date for first-grade enrolment, and report small effects of one year of schooling, net of age/maturity effects. In order to support their interpretation of this difference as reflecting only the different amount of schooling to which children had access on either side of the cut-off, however, it must be assumed that relative age, within a school-entry cohort, does not affect learning; an assumption that is contradicted by empirical observation (Crawford, Dearden and Greaves, 2014[9]). A similar regression-discontinuity analysis (or a fuzzy regression-discontinuity analysis) has since been applied to PISA data for other countries and

economies: to Austria, Croatia and Hungary (Kuzmina and Carnoy, 2016[10]); to Chinese Taipei and Shanghai (China) (Anders, Jerrim and McCulloch, 2016[11]); and to the Russian Federation (Tiumeneva and Kuzmina, 2015[12]). It has also been used to analyse TIMSS 1995 data (an international assessment in which two adjacent grades of primary school students were assessed) (Luyten and Veldkamp, 2011[13]).

This study differs from previous studies based on cross-sectional data in two main ways: first, because it proposes a different identification strategy that does not require strong (and often empirically falsified) assumptions about the effect of being the oldest vs. being the youngest in a school-entry cohort; and second, because, similar to the longitudinal studies cited above, its main thrust is to estimate the joint effect of one year of age and one year of schooling, instead of the "net effect of schooling".

The present paper also contributes to the literature seeking to measure the evolution of test-score gaps across groups of students, as students progress through schooling and enter adult life, either within countries (Bond and Lang, 2013[14]; Fryer and Levitt, 2013[15]; 2004[16]; 2006[17]; 2010[18]; Todd and Wolpin, 2007[19]; Atteberry and McEachin, 2020[20]) or across countries (Singh and Krutikova, 2017[21]; Borgonovi, Choi and Paccagnella, 2021[22]). Most of the latter studies highlight differences in the pace of learning across groups without necessarily quantifying the pace of learning in any of the groups considered. Only a few studies, including the present one, do both, using identical or vertically linked tests; see, for example, Atteberry and McEachin (2020[20]).

Finally, the present paper is also related to more methodological work on the interpretation of test scores and test-score differences. The units of test scores derived from modern educational assessments, such as PISA, do not have a substantive meaning, unlike physical units such as metres or grams. Instead, these units are set in relation to the variation in results observed across all test participants and are computed using item-response-theory models. There is theoretically no minimum or maximum score in PISA; rather, the results are scaled to fit approximately normal distributions (in the case of PISA, the mean is around 500 score points and the standard deviation is around 100 score points). In statistical terms, a one-point difference on the PISA scale therefore corresponds to an effect size (Cohen's d) of 0.01; and a ten-point difference to an effect size of 0.10.

While the units of test scores do not have substantive meaning, a significant body of literature discusses the interpretation of test scores, i.e. the tools that one can use to give test scores a real-world meaning that does not simply refer to a particular test form or to the abstract statistical models used to derive them. For example, Angoff (1984, pp. 44-47[23]) discusses how (empirical) age and grade equivalents, or (normative) age- and grade-specific norms, can be derived and used to interpret test scores. Bloom et al. (2008[24]) illustrate how such age equivalents can be used as benchmarks for interpreting effect sizes from education interventions. Their study also shows that annual grade gains – the combined effect of one year of schooling and of one year of age- decline as students move from the early grades to later grades, irrespective of the subject. The present study contributes to this literature by illustrating how the annual grade gain can be identified in international assessments (and therefore, compared across countries), even in the absence of longitudinal data.

2. Data

The data used in this paper were collected as part of 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 and economies in total, every three years since 2000 (participation of countries/economies has generally increased over time, but not all countries/economies participated in every assessment cycle since they began taking part in PISA). This paper uses, in particular, data for Austria and Scotland from the years 2012, 2015 and 2018. All data are published by the OECD as "public use files" and can be accessed through www.oecd.org/pisa.[1]

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. In addition to test scores, we use variables collected through PISA questionnaires and sampling forms; these include information about students' age, gender, school track, socio-economic status and family background (e.g. immigrant background).[2]

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 paper). PISA participants are selected from the population of 15-year-old students in each country/economy according to a two-stage random sampling procedure. In the first stage, a stratified sample of schools is drawn; in the second stage, students are selected at random in each sampled school. All statistical inference accounts for this complex sample design through resampling methods (replicate weights used to this end are provided with PISA databases).[3]

While PISA data provide a common metric for learning outcomes across education systems that vary significantly in their structure and curricula, they typically are collected over a short data collection period from a single cohort of students (defined by a 12-month window in birthdates). In light of this limited variation, PISA data cannot be readily used to identify the progress that students make from one grade to the next, around the age of 15.

To overcome this limitation, data from multiple editions of PISA (the three most recent cycles: 2012, 2015 and 2018) are used, focussing on Austria and Scotland, the only jurisdictions that changed their testing dates by more than two months over this period.[4] Austria and Scotland share similar levels of mean performance in PISA (Table 2.1 below) but represent two very different models of secondary schooling. Austria is an early-tracking system: after age 10, students are sorted into up to four different school programmes depending on their aptitudes and preferences. In contrast, in Scotland, only a single

[1] The information on students' school track is masked in public use files for Austria in 2018, and was retrieved from restricted-use files.

[2] PISA data include multiple imputations of test scores (plausible values), rather than a single test score variable. In particular, five imputations are included in the 2012 database, and ten imputations are included in later databases. In estimates that combine PISA 2012 data with later data, five imputations are used.

[3] All estimates are computed using the Stata Package repest (Avvisati and Keslair, 2014[41]).

[4] In Annex A, the analysis is extended to 12 more countries or economies that changed their testing dates in earlier years.

education programme is available to students up to the age of 15 (OECD, 2020[25]). The two jurisdictions, therefore, provide an interesting contrast to compare the learning gain over one year of schooling.

2.1. PISA testing dates and samples in Austria and Scotland

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) (OECD, 2017, p. 440[26]). This consistency in testing dates ensures the comparability over time of results, which may otherwise be influenced by 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, this has been the case in Austria and Scotland.

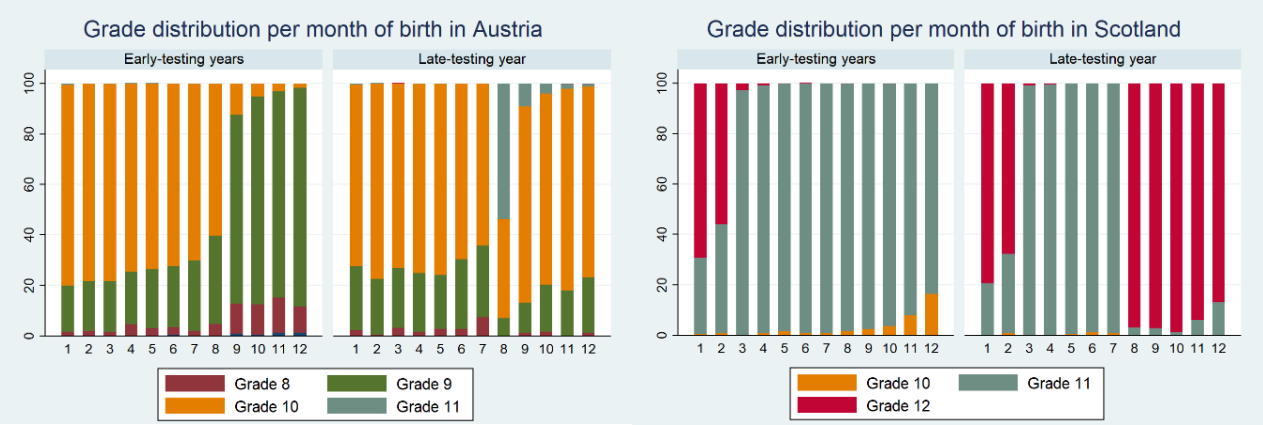
In both Austria and Scotland, the decision to change the testing period was driven mainly by logistic considerations. Austria joined the PISA 2015 cycle late and was therefore allowed to move its testing period to the end of the calendar year; in 2018, the testing dates returned to a period around April (as in earlier cycles). 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 it more closely with that 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.

In both Austria and Scotland, 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 Austria and Scotland), school-entry regulations are centred around a cut-off date that determines eligibility for enrolment in first grade, and define the birthdate of the eldest children in consecutive school-entry cohorts. School-entry regulations in Scotland are such that students born in January and February of a calendar year are expected to start school earlier than students born later. Similar regulations in Austria 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, and between January and February in Scotland, 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 (which is rather frequent in Scotland for students born in January/February), 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 2.1).

It also shows that in the late-testing year (when the PISA cohort was no longer defined by a calendar year), students born between January and July were tested in the same grade level as in early-testing years (but towards the beginning of the school year), while students born between August and December were typically tested in a higher grade level than in early-testing years.

Figure 2.1. Grade distribution per month of birth in Austria and Scotland (United Kingdom)



Note: In Austria, 2012 and 2018 were early-testing years, and 2015 was a late-testing year. In Scotland, 2012 and 2015 were early-testing years, and 2018 was a late-testing year.
Source: PISA 2012, 2015 and 2018 datasets, https://www.oecd.org/pisa/data/ (accessed on 17 May 2021).

Table 2.1 presents descriptive statistics for the samples and main variables of interest used in this paper. Both Austria and Scotland scored, on average, close to the OECD average in the three subjects assessed in PISA (reading, mathematics and science) during the period examined. As expected, given the PISA sampling design, students included in PISA samples were, on average, 15.7 years old regardless of the country and year. They had completed, on average, about 9 grade levels in Austria and about 10.5 grade levels in Scotland since they started primary school (the lower number of completed grades in Austria is due to the later primary school entry and the greater proportion of students behind track). In 2012, about 48% of students in Austria, and about 60% in Scotland, reported that at least one parent had completed a tertiary degree (in both cases, this proportion has been increasing over time, reflecting the increase in educational attainment among parents of 15-year-olds). Throughout the period, in Austria, only about one-third of the PISA cohort was enrolled, at age 15, in a school track with a general academic orientation, with the remaining students attending a pre-vocational or vocational track. In Scotland, the distinction between academic and vocational tracks begins only at a later age.

Table 2.1. Descriptive statistics on PISA samples for Austria and Scotland (United Kingdom)

Year	Variable	Austria					Scotland (United Kingdom)				
		Mean		SD		N	Mean		SD		N
2012	Age (years)[1]	15.7	*(.0)*	0.3	*(.0)*	4 755	15.7	*(.0)*	0.3	*(.0)*	2 945
	Number of completed grade levels (years)[2]	9.0	*(.0)*	0.6	*(.0)*	4 755	10.6	*(.0)*	0.3	*(.0)*	2 945
	General academic track (%)	30.7	*(.9)*			4 755					
	Girl (%)	50.1	*(1.5)*			4 755	49.6	*(1.0)*			2 945
	At least one parent with tertiary-level qual. (%)	48.0	*(1.0)*			4 631	59.6	*(1.1)*			2 798
	Immigrant background (%)	16.5	*(1.1)*			4 695	8.4	*(.8)*			2 867
	Mathematics score	505.5	*(2.7)*	92.5	*(1.7)*	4 755	498.4	*(2.6)*	86.4	*(1.6)*	2 945
	Reading score	489.6	*(2.8)*	91.8	*(1.8)*	4 755	506.1	*(3.0)*	86.7	*(1.8)*	2 945
	Science score	505.8	*(2.7)*	92.2	*(1.6)*	4 755	513.4	*(3.0)*	89.4	*(2.0)*	2 945
2015	Age (years) [1]	15.7	*(.0)*	0.3	*(.0)*	7 007	15.7	*(.0)*	0.3	*(.0)*	3 111
	Number of completed grade levels (years)[2]	8.9	*(.0)*	0.6	*(.0)*	7 006	10.5	*(.0)*	0.3	*(.0)*	3 111
	General academic track (%)	28.6	*(.9)*			7 006					
	Girl (%)	49.5	*(1.5)*			7 007	49.1	*(.6)*			3 111
	At least one parent with tertiary-level qual. (%)	52.7	*(.8)*			6 846	63.9	*(1.1)*			2 853
	Immigrant background (%)	20.3	*(1.1)*			6 928	5.7	*(.5)*			2 956
	Mathematics score	496.5	*(3.0)*	94.8	*(1.9)*	7 007	490.9	*(2.6)*	83.6	*(1.4)*	3 111
	Reading score	484.7	*(2.7)*	101.0	*(1.5)*	7 007	493.3	*(2.1)*	90.5	*(1.5)*	3 111
	Science score	495.0	*(2.5)*	97.1	*(1.3)*	7 007	497.0	*(2.3)*	94.7	*(1.5)*	3 111
2018	Age (years) [1]	15.7	*(.0)*	0.3	*(.0)*	6 802	15.7	*(.0)*	0.3	*(.0)*	2 998
	Number of completed grade levels (years)[2]	8.9	*(.0)*	0.6	*(.0)*	6 802	10.6	*(.0)*	0.5	*(.0)*	2 998
	General academic track (%)	34.2	*(1.1)*			6 802					
	Girl (%)	49.2	*(1.5)*			6 802	50.6	*(.9)*			2 998
	At least one parent with tertiary-level qual. (%)	54.8	*(.7)*			6 611	65.7	*(1.0)*			2 753
	Immigrant background (%)	22.7	*(1.2)*			6 710	8.4	*(.9)*			2 865
	Mathematics score	498.5	*(3.1)*	93.9	*(1.5)*	6 802	488.7	*(4.7)*	93.5	*(2.6)*	2 998
	Reading score	484.2	*(2.7)*	99.3	*(1.2)*	6 802	504.0	*(3.1)*	95.2	*(2.0)*	2 998
	Science score	489.2	*(2.6)*	95.4	*(1.2)*	6 802	490.4	*(3.6)*	96.9	*(2.0)*	2 998

Notes: Means and standard deviations (SD) of Mathematics, Reading and Science scores 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.

1. "Age" is the student's age on a reference date used to determine eligibility for PISA; it is computed based on each student's month and year of birth, rather than from the database variable "age". The latter also accounts for the (limited) variation of testing dates within each PISA sample.

2. The number of completed grade levels is computed as the current grade, minus 1, plus the difference between the age of the student on the reference date (see Note 1) and his or her age at the beginning of the school year (on 1 September).

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

3. Identification strategy

The identification of grade-and-age effects on test performance is usually difficult, if not impossible, in a cross-sectional, single-cohort study such as PISA.

The first challenge is the limited variation in students' age observed among test takers. This implies that strong assumptions are required to identify the effect of one year of age from observed age differences of, at most, 11 months.

The second issue is the fact that the observed variation in grade levels completed 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 birthdate 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).[5] 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.

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 and Scotland, the testing dates observed in PISA begin either in March or October; survey cycles are referred to as "early-testing years" when testing begins in March and as "late-testing years" when testing begins in October.[6] 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

[5] 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[38]; Black, Devereux and Salvanes, 2011[39]; Bedard and Dhuey, 2006[40]). Givord (2020[36]) reviews this literature and provides international evidence based on PISA data.

[6] This notation also allows for a generalisation to other countries, discussed in Annex A.

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 is that grade-and-age differences between late- and early-testing years are negative for some birthdates (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 birthdates (those comprised between August and December, in the case of Austria and Scotland). 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.

However, differences in performance between early- and late-testing years 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. Furthermore, the composition of each cohort may differ, for example, in terms of parental education. There may also be seasonal patterns in test performance or in students' motivation to take a low-stakes test such as PISA.

Yet under the "common-trend" assumption that seasonal patterns of performance and cohort-specific trends are unrelated to a student's month of birth, it is possible to use a double-difference strategy to identify the grade-and-age effect.

Formally, let y_{ist} represent the performance in PISA of student *i*, attending school *s,* in year *t*, 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:

Equation 3.1

$$y_{ist} = \alpha_t + \beta' x_{ist} + \sum_{m=1}^{12} \gamma_m * \mathbb{I}_{m_i=m} + \delta * \mathbb{I}_{t\in L, m_i\in[8,12]} + \epsilon_{is}$$

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 socio-economic status) on performance; ϵ_{ist}, an error term, captures the influence of other student- and school-level characteristics on performance, and $\gamma_{m,t}$ 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 estimated in a flexible way. The only assumption embedded in Equation 3.1 about 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) is that they do not vary over time (other than in ways that are common across all birthdates, captured by α_t). However, when the testing date changes, age at testing and length of schooling change in a discontinuous and discrete way across students' birthdates. 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 in October (late-testing years, $t \in L$), PISA measures the performance of students born between August and December who are one year older and in a higher grade level, compared to students with the same birthdates who would have been eligible for PISA later in the school year, in March. In the equation, the late-testing year is denoted by $t \in L$, where $L = \{2015\}$ in Austria and $L = \{2018\}$ in Scotland.

There are two points worth noting. First, while this double-difference estimator makes it possible to address the identification challenge posed by age-at-entry effects (which would otherwise confound either grade or age effects, or both), it does so at the cost of identifying age and grade effects jointly. Their combined effect is, however, of great interest. It can be directly compared, for example, with estimates from longitudinal studies or from multi-cohort studies. Second, the identification relies on the key assumption that in the absence of changes to the testing schedule, the performance differences across months of birth observed in tests are stable over time. 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).

For example, in the case of Scotland, one must assume that, on average, the same change in score would have been observed between the PISA test in 2015 and 2018 among students born between August and December (who were older and were expected to have completed a few more months of schooling, when the PISA test was conducted in autumn in 2018) as among students born between January and July, had the two groups of students been tested at the same time of the year (in spring, in 2015; and in autumn, in 2018), but had a younger cohort of August-to-December born students sat the test in 2018 (the cohort that was one grade below and one year younger than the one that actually sat the test).

3.1. Assessing the strength of the common-trend assumption

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

The assumption is 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 3.1. 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 suppose 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.

It is not possible to formally test the common-trend assumption, but it is possible to corroborate it with supporting 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 2012 and 2018 in Austria, or between 2012 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 3.1, 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 3.1) such differences to be identical (i.e. to estimate a pseudo difference-in-difference). For example, focussing on students born early in the calendar year only, one could compare those born between January and March to those born between April and June. 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. 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 3.1 using only the observations from 11 out of 12 months (leave-one-out estimator).

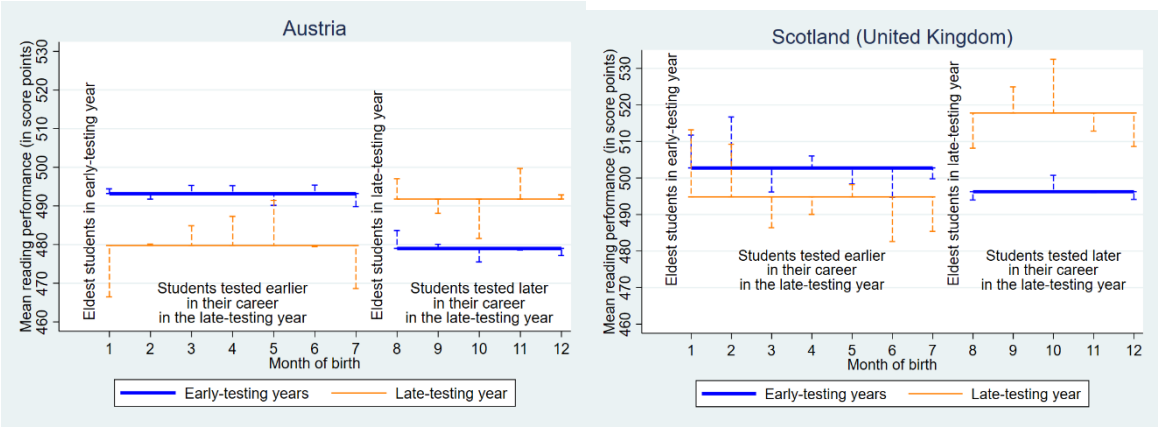
[7] This is similar to applications of difference-in-difference estimators that test the assumption of parallel trends (in the absence of an observed policy change) by showing "underlying" trends prior to the policy change (Angrist and Pischke, 2010, pp. 14-15[37]).

The results of these robustness checks after the main difference-in-difference estimates are presented in the next section.

4. Results

The presentation of the results starts with viewing the graphical intuition behind the difference-in-difference estimator. Figure 4.1 compares the average performance 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 in months such that they are less advanced in their school career (and younger, by a few months) in the year in which testing was conducted later. The second group includes students born in months such that they are more advanced in their school career (and older, by a few months) in the year in which testing was conducted later. 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 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 both Austria and Scotland.

Figure 4.1. Mean performance in reading in Austria and Scotland (United Kingdom), by month of birth and year



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

Table 4.1 shows the average grade-and-age effects estimated by exploiting the variation in testing dates between consecutive PISA cycles (estimates based on Equation 3.1). In the absence of covariates, this corresponds, in Figure 4.1, to the sum of the vertical distances between the two sets of parallel horizontal lines.

The estimates in Table 4.1 imply that students' test scores in PISA increase, over a full school year, on average by about one-fourth of a standard deviation – or around 25 score points – in each subject. The grade gain is more precisely estimated in Austria, compared to Scotland, due to the larger sample size; however, point estimates are similar in both cases, suggesting similar levels of productivity for school education across both systems, in spite of the significant institutional differences (some of these differences are discussed below, in the context of selection effects).

Table 4.1. Grade-and-age effects in Austria and Scotland (United Kingdom)

	Grade-and-age effect[1]						Year fixed effects	Month-of-birth dummies	Covariates[2]	Number of observations
	Mathematics		Reading		Science					
Austria	25.7	*(3.7)*	26.3	*(3.9)*	24.3	*(3.7)*	Yes	Yes	No	18 564
	23.4	*(3.2)*	23.2	*(3.5)*	21.7	*(3.3)*	Yes	Yes	Yes	18 184
Scotland (United Kingdom)	29.7	*(6.6)*	30.2	*(5.5)*	22.8	*(6.5)*	Yes	Yes	No	9 054
	26.6	*(6.6)*	26.2	*(5.7)*	20.1	*(6.8)*	Yes	Yes	Yes	8 522

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. In Austria, 2012 and 2018 were early-testing years, and 2015 was a late-testing year. In Scotland, 2012 and 2015 were early-testing years, and 2018 was a late-testing year.
1. 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 3.1 for details.
2. The following covariates are included: girl, immigrant background, quarter of the index of socio-economic status (three dummies).
Source: PISA 2012, 2015 and 2018 datasets, https://www.oecd.org/pisa/data/ (accessed on 17 May 2021).

4.1. Robustness checks

The interpretation of the double-difference estimator in Table 4.1 as reflecting the grade gain in test scores requires an assumption of parallel changes across groups of students defined by their month of birth. This means that the average difference in performance, compared to early-testing years, which is observed among students born between January and July, would also have been observed among students born between August and December, had they been tested 12 months earlier; or, equivalently, that the average difference in performance observed among students born between August and December in Figure 4.1 would also have been observed among students born between January and July, had they been tested 12 months later. Violations of this hypothesis may result from unique time trends affecting only certain months of birth, or from selection biases that affect differentially one of the two groups.

4.1.1. Parallel trends

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

The first restricts the sample to two assessment years in which the testing occurred on the same dates (2012 and 2015 for Scotland; 2012 and 2018 for Austria). Under the assumption of common trends by month of birth, the interaction term between the 2012 dummy and being born in August through December should not be significant in these regressions. Results, shown in Table 4.2 (Panel A), confirm that this is the case.

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 (students born in January through July). It distinguishes two made-up groups of approximately equal size (January-April vs. May-July). 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 4.2 confirms this to be the case.

Table 4.2. Parallel trends by month of birth in Austria and Scotland (United Kingdom)

Panel A. Are performance changes different for those born in August through December, when the testing date remains the same?

	Placebo effect[1]						Year fixed effects	Month-of-birth dummies	Number of observations
	Mathematics		Reading		Science				
Austria (2012-18)	0.5	*(5.5)*	-5.5	*(5.6)*	0.3	*(5.4)*	Yes	Yes	11 557
Scotland (United Kingdom) (2012-15)	2.6	*(4.6)*	0.4	*(4.7)*	4.1	*(4.8)*	Yes	Yes	6 056

Panel B. Are performance changes in the late-testing year different for students born in January through April, compared to students born in May through July?

	Placebo effect[2]						Year fixed effects	Month-of-birth dummies	Number of observations
	Mathematics		Reading		Science				
Austria	3.0	*(4.3)*	2.4	*(4.4)*	1.1	*(4.3)*	Yes	Yes	10 821
Scotland (United Kingdom)	-0.5	*(6.7)*	-0.8	*(5.8)*	10.0	*(6.8)*	Yes	Yes	5 219

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.
1. 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 2012 and a dummy identifying the months of birth of students who would have been older if tested during a late-testing window; during both years included in the placebo regressions, students were actually tested in the same months. See Equation 3.1 for details.
2. Placebo effects for each subject are estimated using separate regressions. In Panel B, they 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. Only students born in January through July 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. See Equation 3.1 for details.
Source: PISA 2012, 2015 and 2018 datasets, https://www.oecd.org/pisa/data (accessed on 17 May 2021).

4.1.2. Absence of selection effects

The grade-gain estimate for Scotland between the age of 15 and 16 corresponds, for the vast majority of the cohort, to the transition between the grades of Secondary 4 (S4) and Secondary 5 (S5). In the Scottish education system, Secondary 4 marks the end of compulsory schooling, and, for the cohort that participated in PISA 2018, it was possible to leave school education after sitting end of S4 exams (called National 4/5 exams). It is therefore important to investigate whether the composition of the student cohort changed in this transition in ways that could confound the identification of grade gains.

In Austria, the estimated grade gain 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 (OECD, 2020[27]). 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 (Salchegger and Suchań, 2017[28]).

A first indicator of possible selection effects is the proportion of the total population of 15-year-olds represented by PISA samples. While PISA samples are drawn to be representative only of those 15-year-olds who are enrolled in school, this proportion – referred to as "Coverage Index 3" in PISA technical reports (OECD, 2014[29]; 2017[26]; 2020[30]) – can be used to assess to what extent samples from early- and late-testing years are comparable. In both Austria and Scotland, this proportion was a few percentage points lower in the late-testing year than in the early-testing year (see Tables 11.1 and 11.2 in the cited reports). In Austria, Coverage Index 3 dropped from 0.88 in 2012 to 0.83 in 2015 before returning to 0.89 in 2018. In Scotland, it dropped from 0.87 in 2012 and 0.89 in 2015 to 0.85 in 2018. A lower coverage of the total 15-year-old population may reflect a higher proportion of out-of-school youth or of students not listed on the sampling frame, e.g. because of difficulties in collecting complete student lists near the beginning of the school year). If this lower coverage is concentrated among students born in particular months (e.g. the eldest students in the late-testing year), then it may confound the estimates of grade gains in Austria and Scotland. For example, if the lowest-achieving students are more likely to drop out of school, but only after completing Grade 9 in Austria or Secondary 4 in Scotland, then the difference-in-difference estimator will likely overestimate the grade gain.

A test of the presence of selection effects is shown in Table 4.3. 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).[8] The point estimates are close to zero and never 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. While this difference is interpreted 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 tend to be slightly smaller than grade-gain estimates without controls for covariates (Table 4.1).

Table 4.3. Absence of selection bias on grade-gain estimates for Austria and Scotland (United Kingdom)

Difference in sample covariates associated with one additional year of schooling and age

	Selection effect due to age and grade difference						Late-test-window dummy	Month-of-birth dummies	Number of observations
	Girl		Immigrant background		High ESCS[1]				
Austria	0.7	*(1.8)*	-0.4	*(1.6)*	2.9	*(1.9)*	Yes	Yes	18 184
Scotland (United Kingdom)	1.9	*(1.9)*	1.5	*(1.0)*	2.3	*(2.0)*	Yes	Yes	8 522

Notes: Standard errors that account for clustering and for the sampling design are presented in parentheses and italics. 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 **Error! Reference source not found.** Equation 3.1 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.

8. 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.

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

4.1.3. Leave-one-out analysis

The final robustness check consists of estimating Equation 3.1 on 12 different sub-samples, each defined by excluding students born in a particular month from the main sample. If results reported in Table 4.1 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 4.4 shows.

Table 4.4. Robustness of grade-gain estimates in Austria and Scotland (United Kingdom)

	Grade-and-age effect (min-max)[1]						Year fixed effects	Month-of-birth dummies
	Mathematics		Reading		Science			
Austria	23.9	27.9	23.9	28.6	22.5	25.9	Yes	Yes
Scotland (United Kingdom)	27.8	31.1	27.4	32.2	21.6	24.6	Yes	Yes

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. No covariates are included in regressions.
1. 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 3.1 for details.
Source: PISA 2012, 2015 and 2018 datasets, https://www.oecd.org/pisa/data/ (accessed on 17 May 2021).

5. Extensions

In this section, the analysis is extended to explore differences in the grade gain across groups defined by gender, parental background, or (in Austria) the academic or vocational orientation of the study programme.

5.1. Subgroup differences

On average, at age 15 years, boys' performance lags behind girls' performance in reading, but boys score higher than girls in mathematics, 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. In Austria, 15-year-olds who attend school tracks with a general orientation score higher than 15-year-olds who attend vocational or pre-vocational tracks (Table 5.1).

Table 5.1. Differences in mean performance in PISA in Austria and Scotland (United Kingdom), by gender, socio-economic status and school track

	Year	Mathematics			Reading			Science		
Gender gap										
		Boys	Girls	Diff.	Boys	Girls	Diff.	Boys	Girls	Diff.
Austria	2012	517 *(3.9)*	494 *(3.3)*	22.2 *(4.9)*	471 *(4.0)*	508 *(3.4)*	-36.9 *(5.0)*	510 *(3.9)*	501 *(3.4)*	8.6 *(5.0)*
	2015	510 *(3.8)*	483 *(3.8)*	26.7 *(5.0)*	474 *(3.9)*	496 *(3.8)*	-21.7 *(5.5)*	504 *(3.5)*	486 *(3.2)*	18.5 *(4.7)*
	2018	505 *(3.9)*	492 *(4.0)*	13.3 *(5.1)*	470 *(3.6)*	498 *(3.7)*	-27.9 *(5.2)*	490 *(3.7)*	488 *(3.5)*	1.8 *(4.9)*
Scotland (United Kingdom)	2012	506 *(3.0)*	491 *(3.2)*	14.3 *(3.3)*	493 *(3.2)*	520 *(3.5)*	-27.1 *(3.4)*	517 *(3.3)*	510 *(3.6)*	6.7 *(3.3)*
	2015	494 *(2.9)*	487 *(3.6)*	7.1 *(4.0)*	483 *(2.9)*	504 *(2.7)*	-20.8 *(3.8)*	498 *(3.1)*	496 *(2.9)*	1.5 *(3.8)*
	2018	496 *(5.6)*	481 *(4.8)*	14.8 *(4.7)*	496 *(4.0)*	512 *(3.6)*	-15.5 *(4.2)*	494 *(5.1)*	487 *(4.4)*	7.7 *(6.3)*
Socio-economic gap										
		Low ESCS	High ESCS	Diff.	High ESCS	Low ESCS	Diff.	High ESCS	Low ESCS	Diff.
Austria	2012	477 *(3.2)*	536 *(3.1)*	-58.8 *(3.7)*	462 *(3.4)*	520 *(3.3)*	-57.8 *(4.1)*	476 *(3.2)*	538 *(3.0)*	-61.9 *(3.7)*
	2015	468 *(3.2)*	526 *(3.3)*	-58.5 *(3.4)*	454 *(3.3)*	518 *(2.6)*	-64.2 *(3.8)*	463 *(2.7)*	528 *(2.7)*	-64.7 *(3.4)*
	2018	471 *(3.4)*	528 *(3.1)*	-56.8 *(3.4)*	457 *(3.0)*	514 *(2.7)*	-57.1 *(3.3)*	462 *(3.0)*	519 *(2.6)*	-57.9 *(3.3)*
Scotland (United Kingdom)	2012	475 *(3.2)*	525 *(2.8)*	-49.3 *(3.7)*	485 *(3.5)*	530 *(3.3)*	-45.2 *(3.9)*	492 *(3.7)*	539 *(3.3)*	-47.3 *(4.5)*
	2015	468 *(2.6)*	516 *(3.7)*	-48.2 *(3.9)*	472 *(2.8)*	517 *(2.6)*	-44.9 *(3.7)*	472 *(2.7)*	525 *(2.9)*	-52.6 *(3.7)*
	2018	469 *(6.8)*	512 *(4.9)*	-42.5 *(8.6)*	481 *(3.4)*	530 *(4.0)*	-48.4 *(4.9)*	464 *(5.0)*	519 *(4.7)*	-55.0 *(6.8)*
Difference between general and vocational tracks										
		Vocational and pre-vocational	General	Diff.	Vocational and pre-vocational	General	Diff.	Vocational and pre-vocational	General	Diff.
Austria	2012	494 *(3.0)*	532 *(5.9)*	-37.9 *(6.7)*	473 *(3.0)*	527 *(6.0)*	-54.5 *(6.7)*	493 *(2.8)*	535 *(5.7)*	-42.4 *(6.2)*
	2015	480 *(3.4)*	537 *(5.9)*	-56.5 *(6.7)*	463 *(2.9)*	540 *(5.3)*	-77.6 *(6.0)*	475 *(2.7)*	546 *(5.2)*	-71.5 *(5.8)*
	2018	487 *(3.7)*	520 *(4.4)*	-32.5 *(5.2)*	469 *(3.2)*	514 *(4.3)*	-45.3 *(5.0)*	477 *(3.2)*	513 *(4.1)*	-35.7 *(5.0)*

Note: All statistics reported in this table are based on multiply imputed test scores (five plausible values); standard errors that account for clustering and for the sampling design are presented in parentheses and italics.
Source: PISA 2012, 2015 and 2018 datasets, https://www.oecd.org/pisa/data/ (accessed on 17 May 2021).

In order to explore the existence of differential grade gains depending on students' and school characteristics, a modified version of Equation 3.1 is estimated. This version includes additional interaction terms, so that both the underlying trends (represented by α_t), the month-of-birth effects (represented by γ_m) and the grade-gain coefficient δ are allowed to vary across subgroups (the subgroup indicator is also included among the vector of control variables x_{ist}). Table 5.2 reports the jointly estimated grade-and-age effects for

each subgroup, as well as the difference between them. Before commenting on the results, it must be noted that such a triple-difference estimator 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 4.1. Only major differences in the grade gain across subgroups can be detected in the PISA samples used in this analysis.

Table 5.2. Between-group difference in grade-and-age effects in Austria and Scotland (United Kingdom)

	Domain	Grade-and-age effects						Covariates	No. of observations
A. Gender difference		Boys		Girls		Diff.			
Austria	Mathematics	23.2	*(4.6)*	23.5	*(4.5)*	0.2	*(6.3)*	Yes	18 184
	Reading	19.2	*(5.3)*	27.0	*(4.6)*	7.8	*(6.9)*	Yes	18 184
	Science	18.9	*(4.9)*	24.4	*(4.2)*	5.5	*(6.3)*	Yes	18 184
Scotland (United Kingdom)	Mathematics	36.3	*(7.4)*	17.4	*(9.2)*	-18.9	*(10.2)*	Yes	8 522
	Reading	35.0	*(9.0)*	18.2	*(6.6)*	-16.8	*(10.8)*	Yes	8 522
	Science	30.9	*(8.9)*	9.7	*(8.5)*	-21.2	*(10.6)*	Yes	8 522
B. Socio-economic difference		Low ESCS[1]		High ESCS[1]		Diff.			
Austria	Mathematics	26.9	*(5.3)*	20.1	*(4.6)*	-6.8	*(7.5)*	Yes	18 183
	Reading	26.9	*(5.2)*	19.9	*(4.8)*	-7.0	*(7.2)*	Yes	18 184
	Science	24.1	*(4.7)*	19.4	*(4.7)*	-4.7	*(6.8)*	Yes	18 184
Scotland (United Kingdom)	Mathematics	30.2	*(9.8)*	23.4	*(7.2)*	-6.8	*(11.2)*	Yes	
	Reading	35.2	*(7.6)*	17.6	*(7.5)*	-17.6	*(9.7)*	Yes	8 522
	Science	27.3	*(8.4)*	13.5	*(8.2)*	-13.8	*(9.5)*	Yes	8 522
C. Difference between tracks		Vocational and pre-vocational		General		Diff.			
Austria	Mathematics	15.1	*(3.7)*	51.2	*(5.9)*	36.2	*(6.8)*	Yes	18 183
	Reading	16.2	*(4.4)*	51.5	*(5.5)*	35.3	*(7.1)*	Yes	18 183
	Science	13.8	*(4.1)*	49.5	*(5.8)*	35.7	*(7.3)*	Yes	18 183

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. Each row corresponds to a separate regression. Grade-and-age effects for subgroups correspond to the coefficient on the triple interaction term 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 3.1) 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 2012, 2015 and 2018 datasets, https://www.oecd.org/pisa/data/ (accessed on 17 May 2021).

Results show a very large difference in the grade gain between general and vocational tracks in Austria, in favour of the more academically-oriented general tracks attended by about one-third of the cohort (Table 5.2, Panel C). This is consistent with the fact that students in general tracks are exposed, around age 15, to significantly more opportunities to learn the subjects assessed in PISA, compared to students in vocational tracks. Such a difference may also be due to a relative decline among vocational students in their motivation to take the PISA test as they progress further in their studies. A previous study, based on a longitudinal follow-up of PISA student in Germany, documented large declines in test motivation among students in vocational tracks (Heine et al., 2017[31]).

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 (Table 5.2, Panel A). The estimated gender difference is significant at the 5% level in science and at the 10% level in mathematics (and only somewhat smaller in magnitude in reading). The corresponding gender differences in Austria are not significant and close to 0. 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[22]). 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 Austria 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 5.2, Panel B).[9] The negative point estimates suggest that over one year of schooling (and age), the proficiency of disadvantaged children increases 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.

[9] In Scotland, the estimated difference is significant at the 10% level.

6. Discussion

The present paper quantifies the learning gain that results from an additional year of schooling in secondary schools, using data from a large-scale international assessment. Its original identification strategy overcomes the limitations of previous studies that relied on regression-discontinuity designs 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 paper indicate that for 15-year-old students in two distinct European education systems, the typical grade gain is about one-fourth of a standard deviation, or around 25 score points.[10]

The average grade effect for 15-year-olds estimated in the present paper 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 gap between students who are one grade level apart, around the age of 15 (OECD, 2019[32]); as was the gender gap in reading (30 score points, on average across OECD countries). But it would take students in the bottom 25% of socio-economic 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 (OECD, 2019[33]). While tempting, a simple conversion of any difference to years-of-schooling equivalents should, however, be avoided. This would indeed require an extrapolation from the effect of a single grade, around the age of 15, and on average, to the cumulative effect of multiple years of schooling, for a particular group of students – often very different from the group on which the grade gain was estimated.

The present paper also explores the relative importance for skill acquisition of school instruction and of other life experiences. We find, in particular, that the grade gain in Austria differs significantly for students in the general, more academic track, compared to students in vocational tracks.

Previously, this question has mostly been examined in the literature based on seasonal patterns in test scores. 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[34]). 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[35]). 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[20]).

The estimates in this paper suggest that school instruction is the most important factor not only for the mastery of curricular content (which is the focus of most studies on the "summer learning loss") but also for the acquisition of reading, mathematics and science skills measured by PISA, whose tests are not tied to a particular curriculum.

[10] This estimate is remarkably close to the "rule of thumb" that average student learning in a year is equal to about one-quarter to one-third of a standard deviation, suggested in Woessmann (2016[42]).

This implies that periods during which access to school instruction or its quality is greatly reduced for an entire cohort (as has been the case recently, with widespread school closures due to COVID-19) will likely affect the skills of this cohort significantly.

In closing, two important limitations are worth highlighting. A first limitation is that the exogenous source of variation in grade levels for otherwise similar students, which was exploited in the present study, is rare among countries and economies participating in PISA. This means that an estimate of such learning gains for all PISA participants cannot be provided. In Annex A, the corresponding estimates for 12 more countries/economies that changed their testing dates in earlier cycles of PISA are provided, however. A second limitation is related to the fact that PISA data were not designed with the intention of measuring grade-and-age effects; the statistical uncertainty associated with such estimates is therefore relatively large and limits the possibility of analysing differences in the grade gain across subgroups.

Annex A. Grade-and-age effects in 12 additional countries and economies

Table A.1 lists all participants in PISA that changed testing dates since their first participation in PISA, in addition to Austria and Scotland. As can be seen, for most of these countries and economies, the change in testing date occurred in the early cycles of PISA and often coincided with the country's second participation in PISA.

Table A.1. Countries and economies that changed testing dates over the course of their participation in PISA

Country/economy	Last wave before the change in testing date (beginning of testing period)	First wave after the change in testing date (beginning of testing period)
Austria	PISA 2015 (October)[1]	PISA 2018 (March)
Brazil	PISA 2006 (July)	PISA 2009 (March)
Hong Kong (China)	PISA 2000 (December)[2]	PISA 2003 (May)
Indonesia	PISA 2000 (September)[2]	PISA 2003 (April)
Israel	PISA 2000 (January)[2]	PISA 2006 (March)
Macao (China)	PISA 2003 (May)	PISA 2006 (March)
Malaysia	PISA 2009 (June)	PISA 2012 (March)
Mexico	PISA 2000 (February)	PISA 2006 (May)
Romania	PISA 2000 (December)[2]	PISA 2006 (March)
Scotland (United Kingdom)	PISA 2015 (March)	PISA 2018 (October)
Singapore	PISA 2009 (July)	PISA 2012 (March)
Thailand	PISA 2000 (November)[2]	PISA 2003 (August)
United Kingdom (excl. Scotland)[3]	PISA 2003 (March)	PISA 2006 (November)
United States[4]	PISA 2003 (March)	PISA 2003 (September)

Notes:
1. In Austria, the testing date changed in 2015 and reverted to the "typical" testing date for Austria from earlier assessments in 2018.
2. Hong Kong (China), Indonesia and Thailand conducted the PISA 2000 assessment in 2001; Israel and Romania conducted the assessment in 2002, as part of PISA 2000+.
3. School participation rates in the United Kingdom in 2003 were below PISA standards. Identification of grade-and-age effects based Equation 3.1 rests on the assumption that the resulting non-response bias is unrelated to students' month of birth.
4. In the United States, PISA 2003 data were collected in two batches, due to low response rates achieved during the originally planned testing window. Non-responding schools were contacted again in the fall of 2003 to complete the assessment and meet response-rate standards. A different 12-month cohort of eligible students was defined for these late-testing schools. Identification of grade-and-age effects based on Equation 3.1 rests on the assumption that the selection bias in each batch of data is unrelated to students' month of birth.
Source: PISA 2000-12 datasets, https://www.oecd.org/pisa/data/ (accessed on 17 May 2021).

This annex reports estimates of grade-and-age effects for an additional 12 participants in PISA, based on an appropriately modified version of Equation 3.1. In particular, for each participant, the months of birth of the eldest students in "early" and "late" testing years may differ from January and August, the corresponding months for Austria and Scotland. The indicator variable that identifies the two groups of students whose results are compared across PISA cycles is redefined accordingly.

Table A.2 shows the average grade effects estimated in these 12 countries and economies.

Table A.2. Grade-and-age effects in 12 additional countries and economies

Country/economy and years	Grade-and-age effect						Year fixed effects	Month-of-birth dummies	Covariates	Number of observations
	Mathematics		Reading		Science					
Brazil (2006-09)	12.5	*(4.5)*	14.6	*(4.6)*	10.2	*(4.2)*	Yes	Yes	No	29 422
Hong Kong (China) (2001-03)			25.6	*(3.6)*			Yes	Yes	No	8 883
Indonesia (2001-03)			12.4	*(5.3)*			Yes	Yes	No	18 129
Israel (2002-06)			8.6	*(9.0)*			Yes	Yes	No	9 082
Macao (China) (2003-06)	20.2	*(10.7)*	18.1	*(6.6)*			Yes	Yes	No	6 010
Mexico (2000-06)			16.2	*(8.8)*			Yes	Yes	No	34 582
Malaysia (2009-12)	10.2	*(4.3)*	10.2	*(4.5)*	5.8	*(4.0)*	Yes	Yes	No	10 196
Romania (2002-06)			10.0	*(8.4)*			Yes	Yes	No	9 947
Singapore (2009-12)	24.1	*(3.7)*	18.0	*(3.4)*	19.6	*(3.6)*	Yes	Yes	No	10 829
Thailand (2001-03)			13.5	*(5.7)*			Yes	Yes	No	10 576
United Kingdom (excl. Scotland) (2003-06)	19.3	*(4.5)*	30.9	*(5.5)*			Yes	Yes	No	17 037
United States (2003)	20.0	*(7.0)*	15.4	*(8.0)*			Yes	Yes	No	5 306

Notes: 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 3.1 for details. Grade-and-age effects are estimated only for subjects whose reporting scales are equivalent to those currently in use for all years involved in the analysis. In particular, the current reporting scale for science was established only in PISA 2006, and the current reporting scale for mathematics only in PISA 2003. 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.
Source: PISA 2000-12 datasets, https://www.oecd.org/pisa/data/ (accessed on 17 May 2021).

For countries and economies whose mean scores are close to, or above, the OECD average, such as Hong Kong (China), Macao (China), Singapore, the United Kingdom and the United States, the point estimate for the grade gain ranges between 15 score points (reading, in the United States) and 31 score points (reading, in the United Kingdom). These estimates are relatively close to those for Austria and Scotland in more recent years; it must be noted that the confidence intervals around these estimates can be wide. In contrast, in the remaining, lower-performing countries, the point estimates for the grade gain tend to be smaller – between 6 score points (science, in Malaysia) and 16 score points (reading, in Mexico).

References

Anders, J., J. Jerrim and A. McCulloch (2016), "How Much Progress Do Children in Shanghai Make Over One Academic Year? Evidence From PISA", *AERA Open*, Vol. 2/4, p. 233285841667884, http://dx.doi.org/10.1177/2332858416678841. [11]

Andrabi, T. et al. (2011), "Do Value-Added Estimates Add Value? Accounting for Learning Dynamics", *American Economic Journal: Applied Economics*, Vol. 3/3, pp. 29-54, http://dx.doi.org/10.1257/app.3.3.29. [5]

Angoff, W. (1984), *Scales, norms, and equivalent scores*, Educational Testing Service, https://www.ets.org/Media/Research/pdf/Angoff.Scales.Norms.Equiv.Scores.pdf. [23]

Angrist, J. and J. Pischke (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics", *Journal of Economic Perspectives*, Vol. 24/2, pp. 3-30, http://dx.doi.org/10.1257/jep.24.2.3. [37]

Atteberry, A. and A. McEachin (2020), "School's Out: The Role of Summers in Understanding Achievement Disparities", *American Educational Research Journal*, p. 000283122093728, http://dx.doi.org/10.3102/0002831220937285. [20]

Avvisati, F. and F. Keslair (2014), *REPEST: Stata module to run estimations with weighted replicate samples and plausible values*, Statistical Software Components S457918, Boston College Department of Economics. [41]

Bedard, K. and E. Dhuey (2006), "The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects", *The Quarterly Journal of Economics*, Vol. 121/4, pp. 1437-1472, http://dx.doi.org/10.1093/qje/121.4.1437. [40]

Black, S., P. Devereux and K. Salvanes (2011), "Too Young to Leave the Nest? The Effects of School Starting Age", *Review of Economics and Statistics*, Vol. 93/2, pp. 455-467, http://dx.doi.org/10.1162/rest_a_00081. [39]

Bloom, H. et al. (2008), "Performance Trajectories and Performance Gaps as Achievement Effect-Size Benchmarks for Educational Interventions", *Journal of Research on Educational Effectiveness*, Vol. 1/4, pp. 289-328, http://dx.doi.org/10.1080/19345740802400072. [24]

Bond, T. and K. Lang (2013), "The Evolution of the Black-White Test Score Gap in Grades K–3: The Fragility of Results", *Review of Economics and Statistics*, Vol. 95/5, pp. 1468-1479, http://dx.doi.org/10.1162/rest_a_00370. [14]

Borgonovi, F., A. Choi and M. Paccagnella (2021), "The evolution of gender gaps in numeracy and literacy between childhood and young adulthood", *Economics of Education Review*, Vol. 82, p. 102119, http://dx.doi.org/10.1016/j.econedurev.2021.102119. [22]

Chetty, R., J. Friedman and J. Rockoff (2014), "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates", *American Economic Review*, Vol. 104/9, pp. 2593-2632, http://dx.doi.org/10.1257/aer.104.9.2593. [6]

Cooper, H. et al. (1996), "The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic Review", *Review of Educational Research*, Vol. 66/3, pp. 227-268, http://dx.doi.org/10.3102/00346543066003227. [34]

Crawford, C., L. Dearden and E. 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)*, Vol. 177/4, pp. 829-860, http://dx.doi.org/10.1111/rssa.12071. [9]

Dearden, L., C. Crawford and C. Meghir (2010), "When you are born matters: the impact of date of birth on educational outcomes in England", *Working Paper Series*, Institute for Fiscal Studies, http://dx.doi.org/10.1920/wp.ifs.2010.1006. [38]

Fryer, R. and S. Levitt (2013), "Testing for Racial Differences in the Mental Ability of Young Children", *American Economic Review*, Vol. 103/2, pp. 981-1005, http://dx.doi.org/10.1257/aer.103.2.981. [15]

Fryer, R. and S. Levitt (2010), "An Empirical Analysis of the Gender Gap in Mathematics", *American Economic Journal: Applied Economics*, Vol. 2/2, pp. 210-240, http://dx.doi.org/10.1257/app.2.2.210. [18]

Fryer, R. and S. Levitt (2006), "The Black-White Test Score Gap Through Third Grade", *American Law and Economics Review*, Vol. 8/2, pp. 249-281, http://dx.doi.org/10.1093/aler/ahl003. [17]

Fryer, R. and S. Levitt (2004), "Understanding the Black-White Test Score Gap in the First Two Years of School", *Review of Economics and Statistics*, Vol. 86/2, pp. 447-464, http://dx.doi.org/10.1162/003465304323031049. [16]

Givord, P. (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, https://dx.doi.org/10.1787/822ea6ce-en. [36]

Heine, J. et al. (2017), "Empirische Grundlage, Stichprobenausfall und Adjustierung im PISA-Längsschnitt 2012–2013", *Zeitschrift für Erziehungswissenschaft*, Vol. 20/S2, pp. 287-306, http://dx.doi.org/10.1007/s11618-017-0756-0. [31]

Jones, S. et al. (2014), "Can Your Child Read and Count? Measuring Learning Outcomes in East Africa", *Journal of African Economies*, Vol. 23/5, pp. 643-672, http://dx.doi.org/10.1093/jae/eju009. [2]

Kane, T. and D. Staiger (2008), *Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation*, National Bureau of Economic Research, Cambridge, MA, http://dx.doi.org/10.3386/w14607. [7]

Kuzmina, J. and M. Carnoy (2016), "The effectiveness of vocational versus general secondary education", *International Journal of Manpower*, Vol. 37/1, pp. 2-24, http://dx.doi.org/10.1108/ijm-01-2015-0022. [10]

Luyten, H., J. Peschar and R. Coe (2008), "Effects of Schooling on Reading Performance, Reading Engagement, and Reading Activities of 15-Year-Olds in England", *American Educational Research Journal*, Vol. 45/2, pp. 319-342, http://dx.doi.org/10.3102/0002831207313345. [8]

Luyten, H. and B. 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*, Vol. 4/3, pp. 264-288, http://dx.doi.org/10.1080/19345747.2010.519825. [13]

Nagy, G. et al. (2017), "IRT-Skalierung der Tests im PISA-Längsschnitt 2012/2013: Auswirkungen von Testkontexteffekten auf die Zuwachsschätzung", *Zeitschrift für Erziehungswissenschaft*, Vol. 20/S2, pp. 229-258, http://dx.doi.org/10.1007/s11618-017-0749-z. [4]

OECD (2020), "Characteristics of education systems", in *Education at a Glance 2020: OECD Indicators*, OECD Publishing, Paris, https://dx.doi.org/10.1787/4bb100a8-en. [27]

OECD (2020), *PISA 2018 Technical Report*, https://www.oecd.org/pisa/data/pisa2018technicalreport/. [30]

OECD (2020), "Sorting and selecting students between and within schools", in *PISA 2018 Results (Volume V): Effective Policies, Successful Schools*, OECD Publishing, Paris, https://dx.doi.org/10.1787/5d9b15a4-en. [25]

OECD (2019), *PISA 2018 Results (Volume I): What Students Know and Can Do*, OECD Publishing. [32]

OECD (2019), *PISA 2018 Results (Volume II): Where All Students Can Succeed*, PISA, OECD Publishing, Paris, https://dx.doi.org/10.1787/b5fd1b8f-en. [33]

OECD (2017), *PISA 2015 Technical Report*, http://www.oecd.org/pisa/data/2015-technical-report/ (accessed on 31 July 2017). [26]

OECD (2014), *PISA 2012 Technical Report*, OECD Publishing, https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf. [29]

Prenzel, M. et al. (eds.) (2006), *PISA 2003: Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*, Waxmann Verlag GmbH. [3]

Salchegger, S. and B. Suchań (2017), "Was bedeutet es für den Geschlechterunterschied in der Mathematikkompetenz bei PISA, wenn dem Schulsystem leistungsschwache Jungen verloren gehen?", *Zeitschrift für Bildungsforschung*, Vol. 8/1, pp. 81-99, http://dx.doi.org/10.1007/s35834-017-0190-7. [28]

Singh, A. (2019), "Learning More with Every Year: School Year Productivity and International Learning Divergence", *Journal of the European Economic Association*, Vol. 18/4, pp. 1770-1813, http://dx.doi.org/10.1093/jeea/jvz033. [1]

Singh, A. and S. Krutikova (2017), "Starting together, growing apart: gender gaps in learning from preschool to adulthood in four developing countries", *Young Lives Working Papers*, Young Lives, https://ora.ox.ac.uk/objects/uuid:c9f322a6-f320-4206-9994-84a7ff0d7479. [21]

Tiumeneva, Y. and J. Kuzmina (2015), "The Difference That One Year of Schooling Makes for Russian Schoolchildren", *Russian Education & Society*, Vol. 57/4, pp. 214-253, http://dx.doi.org/10.1080/10609393.2015.1068567. [12]

Todd, P. and K. Wolpin (2007), "The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps", *Journal of Human Capital*, Vol. 1/1, pp. 91-136, http://dx.doi.org/10.1086/526401. [19]

von Hippel, P. and C. 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*, Vol. 6, pp. 43-80, http://dx.doi.org/10.15195/V6.A3. [35]

Woessmann, L. (2016), "The Importance of School Systems: Evidence from International Differences in Student Achievement", *Journal of Economic Perspectives*, Vol. 30/3, pp. 3-32, http://dx.doi.org/10.1257/jep.30.3.3. [42]