Organisation for Economic Co-operation and Development

EDU/WKP(2021)14

Unclassified **English text only**

18 October 2021

DIRECTORATE FOR EDUCATION AND SKILLS

How much do 15-year-olds learn over one year of schooling? An international comparison based on PISA

OECD Education Working Paper No. 257

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, edu.pisa@oecd.org

JT03483202

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, Francesca Borgonovi, Tiago Fragoso, Pierre Gouëdard, Miyako Ikeda and Marco Paccagnella. The authors also thank Julie Harris for editing the paper and Thomas Marwood for his help in formatting and publishing the paper.

Abstract

This paper quantifies the learning gain that accrues to 15-year-old students over one year of schooling in 18 countries and economies, where the cohort eligible to sit the OECD Programme for International Student Assessment (PISA) test overlaps with two distinct school cohorts. School-entry regulations are used as an exogenous source of variation for grade levels in an instrumental-variables framework. The focus on the joint effect of schooling and age, together with (local) linearity assumptions, make it possible to account for differences in school-starting age across students who are expected to be in different grades. On average, students' test scores increase by about one-fifth of a standard deviation over a school year. While estimates of the grade gain for individual countries and economies come with wide confidence intervals, this study also shows the annual learning gain of students around the age of 15 tends to be larger in high-income countries compared to middle-income countries.

Table of contents

Tables

Figures

1. Introduction

When countries were hit by the public health crisis caused by the novel coronavirus (COVID-19) in early 2020, all but a minority reacted by closing schools for extended periods of time. Schools and education systems were often quick to organise remote support for home-based learning, but several observers have questioned the effectiveness of these schooling surrogates, either in general or for particular types of students. First data from national assessments conducted in 2020 or 2021 confirm, indeed, that the results of many students (particularly those from disadvantaged backgrounds) lagged behind the results of similar students in previous school years (Rose et al., 2021[1]; Renaissance Learning, Education Policy Institute, 2021[2]; Engzell, Frey and Verhagen, 2021[3]; Kuhfeld et al., 2020[4]; Maldonado and De Witte, 2020[5]; INVALSI, 2021[6]).

Such learning losses are often compared to the typical learning progression observed in normal times over a year of schooling. This paper quantifies this "normal" progression in terms of test scores from the OECD Programme for International Student Assessment (PISA). The results can then be used as a benchmark for other differences in performance, such as gender gaps, socio-economic gaps, between-country differences or, indeed, the differences that will eventually be observed between the pre-COVID assessments (PISA 2018) and the post-COVID assessments (PISA 2022).

In order to quantify average learning progressions in PISA, several challenges need to be overcome: PISA tests are administered to a single cohort; there is no longitudinal follow-up measure; and the current grade level attended by 15-year-olds may depend on unobserved determinants such as prior performance or students' health, which also affect the test score in PISA (in other words, the observed variation in grades is, at least in part, endogenous).[1] This paper tries to address these challenges by modelling the relationship between students' test scores, their grade, age at school entry and age on the testing day; and by exploiting the exogenous source of variation in students' grade and age at school entry resulting from school-entry regulations.

This study shows that, on average across 18 countries and economies, students' learning progress around the age of 15 is equivalent to about one-fifth of a standard deviation in students' test scores per school year. While estimates of the grade gain for individual countries and economies come with wide confidence intervals, this study also shows that there is significant variation in the learning gain of students across countries and economies, and that the annual learning gain of students around the age of 15 tends to be larger in high-income countries compared to middle-income countries.

In Section 2, prior studies on "grade effects" based on PISA (or similar data) are reviewed and critiqued, highlighting differences with the present paper. Section 3 describes the data and samples used. Section 4 presents the identification strategy, i.e. the regression function that is estimated and the assumptions that allow for the interpretation of one of the estimated parameters as the causal effect of an additional year of schooling on learning. Section 5 presents the results, and Section 6 compares results across countries and in relation to other published results.

[1] This endogeneity means that naïve comparisons of student performance across grades, as reported in Table A1.2 in past PISA reports (OECD, 2013[37]; 2010[38]; 2007[39]), cannot be interpreted as reflecting causal associations, even when they account for observable differences across students enrolled in different grades.

2. Related literature

Several authors have previously used cross-sectional data from PISA (as in the present study) or from other international large-scale assessments to quantify the effect of an additional grade level on test performance. Most of these studies focus on pure length-of-schooling effects (net of age differences typically associated with grade progressions), and exploit the discontinuity in the age-to-grade mapping due to school-entry regulations in order to address the (partly) endogenous sorting of students across grades.

Luyten, Peschar and Coe (2008[7]) compare, using PISA 2000 data from the United Kingdom (excluding Scotland), students who were born just before and immediately after the cut-off date for first-grade enrolment, and report small effects (equivalent to 12% of a standard deviation) of one year of schooling, net of age/maturity effects. A similar regression-discontinuity analysis (or a fuzzy regression-discontinuity analysis) has since been applied to PISA data for other countries. These include Austria, Croatia and Hungary (Kuzmina and Carnoy, 2016[8]);[2] Chinese Taipei and Shanghai (China) (Anders, Jerrim and McCulloch, 2016[9]);[3] and the Russian Federation (Tiumeneva and Kuzmina, 2015[10]).[4] It has also been used to analyse Trends in International Mathematics and Science Study (TIMSS) 1995 data (an international assessment in which two adjacent grades of primary school students were assessed) (Luyten and Veldkamp, 2011[11]).

In order to support the interpretation test-score differences between children born "just before" and "just after" the cut-off date for first-grade enrolment as reflecting only the different amount of schooling to which children on either side of the cut-off had access, however, a critical assumption is that relative age, within a school-entry cohort, does not affect learning. Yet this assumption is, in many school systems, contradicted by empirical observation (Crawford, Dearden and Greaves, 2014[12]; Black, Devereux and Salvanes, 2011[13]; Dhuey, 2016[14]). As a result, some of the small effects reported by regression-discontinuity studies may reflect the joint effect of one additional year of schooling and of being the youngest (rather than the oldest) child in the class. Some authors, moreover, have pointed out that birth dates may sometimes be chosen (e.g. in the case of scheduled deliveries) (Kim, 2021[15]; Neugart and Ohlsson, 2012[16]; Gans and Leigh, 2009[17]; Huang, Zhang and Zhao, 2020[18]), and different types of families may select birth dates on either side of the cut-off; if the resulting variation in family characteristics on either side of the cut-off is related to schooling outcomes, the effect

[2] Kuzmina and Carnoy (2016, pp. 18-19, Table X[8]) report that a year of schooling increases math scores by 14-16 points (about 0.15 standard deviations) in vocational secondary education and 14-27 points in general education (about 0.15-0.3 standard deviations) in Austria, Croatia and Hungary. It is unclear whether their regression model also accounts for age differences across grades or whether such age differences contribute to the reported effects of a year of schooling.

[3] Anders, Jerrim and McCulloch (2016, pp. 6-7, Table 2[9]) report that the first year of upper secondary schooling in Shanghai leads to an increase of 12 points (0.12 standard deviations) in mathematics, 8.5 points (0.085 standard deviations) in reading and 2.9 points (0.03 standard deviations) in science scores, and to even smaller score gains in Chinese Taipei. Most of these estimates are not statistically significant at conventional levels.

[4] Tiumeneva and Kuzmina (2015, pp. 248-250, Table A2[10]) report point estimates for grade effects in the Russian Federation that are around 15 score points. Due to large standard errors, however, these estimates are not statistically significant. It is unclear whether their regression model also accounts for age differences across grades or whether such age differences contribute to the reported effects of a year of schooling.

of these family characteristics also contributes to the difference in test scores around the cut-off, confounding its interpretation as a length-of-schooling effect.

A different strategy, and one that avoids strong assumptions about month-of-birth or school-starting-age effects, has been used by Avvisati and Givord (2021[19]). The authors exploit the change, across different cycles, in the time of the year when PISA was conducted and estimate the joint effect of one year of schooling and one year of age by comparing, within each country, the PISA scores of students born in the same calendar month across survey cycles that differ in terms of testing dates. Results for Austria and Scotland (United Kingdom), which refer to the 2012-18 period, indicate a grade gain of about 25 score points in both countries and for all three subjects (mathematics, reading and science). Results for 12 additional countries and economies, provided in Annex A (Avvisati and Givord, 2021[19]) and which refer to earlier periods, suggest that in some countries, such as Brazil, Indonesia, Israel, Malaysia, Romania and Thailand, the grade gain corresponds to only about 10-15 score points.

Like the studies mentioned earlier, the present paper exploits the variation in grade levels within the PISA cohort induced by school-entry regulations as a source of identification; but, like Avvisati and Givord (2021[19]), it focuses on the joint effect of schooling and age, rather than on pure length-of-schooling effects. Indeed, rather than comparing students born “just before” and “just after” the cut-off date, it compares the eldest and youngest students in the PISA cohort – which are both (almost) one year of age apart and expected to be one grade level apart.

There are two main reasons that justify the focus of this paper on the joint effect of age and schooling rather than on the pure effect of schooling. The first reason is a substantive reason: the average grade gain – the joint effect of age and schooling – is a more intuitive benchmark for comparing differences in PISA scores because it relates to the typical progression in age and grades that can be observed over time for individual students. In the absence of calendar-year effects, the joint effect of age and schooling corresponds to the difference observed in longitudinal studies in which the same students are followed from one year to the next (Singh, 2019[20]; Prenzel et al., 2006[21]).[5] In the absence of cohort effects, it also corresponds to the difference between the test scores of students in adjacent cohorts or grades, in studies that administer the same test to multiple grades or cohorts (Avvisati and Viltard, forthcoming[22]). It is possible to interpret this joint effect as an average treatment effect of a very common “treatment” and to explicitly acknowledge individual variation in the gains underlying this average (Rubin, 1974[23]). In contrast, it is uncommon to observe variations in an individual’s length of schooling without a corresponding variation in an individual’s age, outside of situations in which some students are prevented from attending school.

The second reason is that such joint effects are identified under less restrictive assumptions about school-starting-age effects, compared to pure length-of-schooling effects. Indeed, in countries where the school-entry cohort and the PISA cohort do not coincide, the eldest and the youngest student in PISA have (almost) the same school-starting age and differ only in terms of age and length of schooling; it is therefore not necessary to assume that school-starting age is exactly 0 in order to identify the joint effect of age and schooling. Furthermore, in contrast to studies exploiting the discontinuity in the age-to-grade mapping around the school-entry cut-off date, the identification of the combined effect of schooling and age, which exploits the discontinuity in the school-starting-age-to-grade mapping, is possible even in the presence of birth-date manipulation around the cut-off date for

[5] To interpret differences in test scores over time as reflecting age- and schooling-effects, test administration conditions must remain stable.

primary school entry. Indeed, it is possible to restrict the sample to exclude those birth dates around the cut-off for school entry, which are most prone to manipulation.

3. Data

All data used in this paper are based on datasets collected in 2015 and 2018 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 has generally increased over time, but not all countries participated in every assessment cycle since they began taking part in 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. In addition to test scores, variables collected through PISA questionnaires and sampling forms are used; these include information about students' age, gender, socio-economic status and family background (e.g. immigrant background).

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 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).[6]

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 birth dates). Because 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.

3.1. Selection of countries

In order to overcome this limitation, the analyses in this paper are limited to a set of countries in which the PISA cohort is expected to be found in two distinct grades. In general, the expected grade of the PISA cohort depends on two elements: the local regulations regarding the birth dates eligible for enrolment in first grade (typically summarised in a cut-off date, with students who turn a certain age by that date being eligible) and the testing date in PISA. For example, in a country whose testing window started in March 2018, all students born in 2002 were considered eligible for PISA. Many countries and economies that participate in PISA choose the testing date so that the PISA cohort coincides with a single school-entry cohort. However, if in a given country students had to turn six by 31 July 2008 in order to enrol in Grade 1 in August or September 2008, one can expect most students born between August and December to be in Grade 9 at the time of testing, and most students born between January and July 2002 to be in Grade 10.

[6] All estimates are computed using the Stata Package repest (Avvisati and Keslair, 2014[48]).

The selection of countries and economies is based on a database containing country-level information on school-entry regulations, which is used to define a variable with the "expected grade" for each month of birth (and, therefore, for each PISA student). This database builds on the responses of country representatives to a system-level questionnaire (Givord, 2020[24]).[7]

The first selection of countries and economies for this study includes more than 20 in which the cut-off for PISA eligibility differs from the cut-off for first-grade enrolment, and, therefore, the expected grade varies within the PISA cohort. The countries and economies included in this study are further restricted to those for which a regression of the actual grade on age (in months) and on the theoretical grade (as determined by the cut-off date) indicates that the cut-off date acts as a binding constraint, and is positively related to the actual grade. This results in the exclusion of eight countries with very open enrolment policies (such as Ireland) or whose cut-off date may have been misreported.[8] The United Kingdom is also excluded from the analysis due to stark differences in cohort coverage rates by month of birth.[9] Finally, Scotland and Austria, which moved their testing dates between 2015 and 2018, are excluded from the analysis. Data for Scotland and Austria are analysed (under less restrictive assumptions) in a companion paper (Avvisati and Givord, 2021[19]).

The final selection of countries and economies includes 18 PISA participants: Albania, Baku (Azerbaijan), Beijing-Shanghai-Jiangsu-Zhejiang (China) (hereafter "B-J-S-Z [China]"), Belarus, Costa Rica, Croatia, the Czech Republic, Estonia, Finland, Germany, Hong Kong (China), Hungary, Korea, Luxembourg, Serbia, the Slovak Republic,

[7] Some corrections are made to the original database using information available on official websites or in prior research. First, three countries are excluded where a cut-off date was reported (Chile, Peru and the United Arab Emirates). In Chile, the actual cut-off date for school enrolment varies locally and may coincide with the month of birth of the eldest students in PISA (the cut-off month for the PISA cohort) (Cáceres-Delpiano and Giolito, 2018[40]). In Peru and the United Arab Emirates, the reported cut-off date cannot plausibly explain the observed grade distribution. Second, in federal countries where cut-off dates vary locally, both the earliest and the latest cut-off date (between 30 June and 30 September in Germany; between 31 March and 31 July in Switzerland) are considered. Third, information are added for the Chinese provinces participating in PISA and for Costa Rica whose cut-off dates for enrolment in first grade – 31 August in China (Liu and Li, 2016[42]; Zhang and Xie, 2018[43]); 15 November in Costa Rica (Fernández Aráuz, 2016, p. 53[44]) – were missing in the original database. Fourth, the cut-off date for the Czech Republic was corrected, based on correspondence with representatives of the Czech School Inspectorate.

[8] The following are excluded, in particular: Bosnia and Herzegovina, Malaysia, Moldova, Portugal, Qatar, Turkey and Uruguay, where a higher expected grade appears to increase the probability of being in a higher grade by less than 10 percentage points (and sometimes appears to negatively relate to that probability).

[9] In the United Kingdom (excluding Scotland and Northern Ireland), the expected grade for the PISA cohort varies only in PISA 2018 data; PISA 2015 was conducted about one month later in the year, and the corresponding assessment cohort coincided with the school-entry cohort in first grade. In 2018, the variation concerns only students born in August. Further analyses showed that the coverage rate for students born in August (and enrolled in a higher grade) was significantly lower than for students born in the remaining months. For example, there are only 367 students in the PISA public-use files for the United Kingdom (excluding Scotland) who were born in August, compared to 959 students born in July and 978 students born in September (public-use files include only responding students). Any comparison of students born in August to students born in the remaining months may therefore be confounded by selection effects driven by non-coverage. Similar analyses were conducted for all participants where the expected variation in expected grades was limited to only one or two months.

Switzerland and Chinese Taipei. Table 3.1 shows the PISA cohorts and the cut-off dates for these countries and economies.

In general, all PISA 2015 and PISA 2018 data are used in this analysis when both years are available. In some countries, however, the sample of students included in the analysis is further restricted.

Table 3.1. School-entry regulations, PISA cohorts and samples used in the analysis

Country/economy and years	PISA cohort: Month of birth of the eldest students	Cut-off date for first grade enrolment[1]	Expected grade level, in PISA, of youngest/eldest students	Birth dates (months) excluded from the analysis	Final sample size after exclusions
Albania (2018)	January	1 September	9/10		6 359
Baku (Azerbaijan) (2018)	January	15 September	9/10	September	6 202
Belarus (2018)	January	1 September	9/10		5 803
B-S-J-Z (China) (2018)	January	1 September	9/10		12 058
Costa Rica (2015, 2018)	March	15 November	9/10	November	14 087
Croatia (2015, 2018)	January	1 April	9/10		12 418
Czech Republic (2015, 2018)	January	1 September	9/10	June-August	13 913
Estonia (2015, 2018)	January	1 October	8/9		10 903
Finland (2015, 2018)	February	1 January	8/9		11 531
Germany[2] (2015, 2018)	January	1 October	9/10	June-September	11 955
Hong Kong (China) (2015, 2018)	February	1 January	9/10		11 396
Hungary (2015, 2018)	January	1 June	8/9		10 790
Korea (2018)	March	1 January	9/10		6 650
Luxembourg (2015, 2018)	January	1 September	9/10		10 529
Serbia (2018)[3]	January	1 March	9/10		6 596
Slovak Republic (2015, 2018)	January	1 September	9/10	June-August	12 315
Switzerland[2] (2015, 2018)	January	1 August	9/10	April-July	7 823
Chinese Taipei (2015, 2018)	January	1 September	9/10		14 951

Notes:
1. Where the cut-off date was reported as the last day in a month, the first day of the following month is reported in this table for consistency across countries and economies.
2. The cut-off date varies locally.
3. Thirteen observations are dropped from the PISA samples to restrict the PISA cohort to a 12-month window of birth dates.
Source: OECD PISA 2015 and 2018 datasets, https://www.oecd.org/pisa/data/ (accessed on 17 May 2021).

For two countries, the estimation sample does not include PISA 2015 data: Albania (because key control variables are not available) and Korea (because the cut-off date for primary school enrolment differed across the two cohorts).[10]

For five countries, it excludes students born in certain months. As is explained below, a critical assumption to interpret the estimates as reflecting causal effects is that the effects of being older, at the time of the test (age-at-the-test effect) and at school start (school-starting-age effect) are approximately linear. This assumption is violated, for example, when primary school enrolment rules foresee an exception for students born

[10] The cut-off date for the PISA 2015 cohort in Korea was 1 March; Korea changed the cut-off date for first-grade enrolment in 2010 (with 2009 as a transition year, in which only students born between March and December 2002 entered primary school) (Kim, 2021[15]). This leads to the exclusion of the PISA 2015 cohort (which coincides with a school-entry cohort) from the estimation samples.

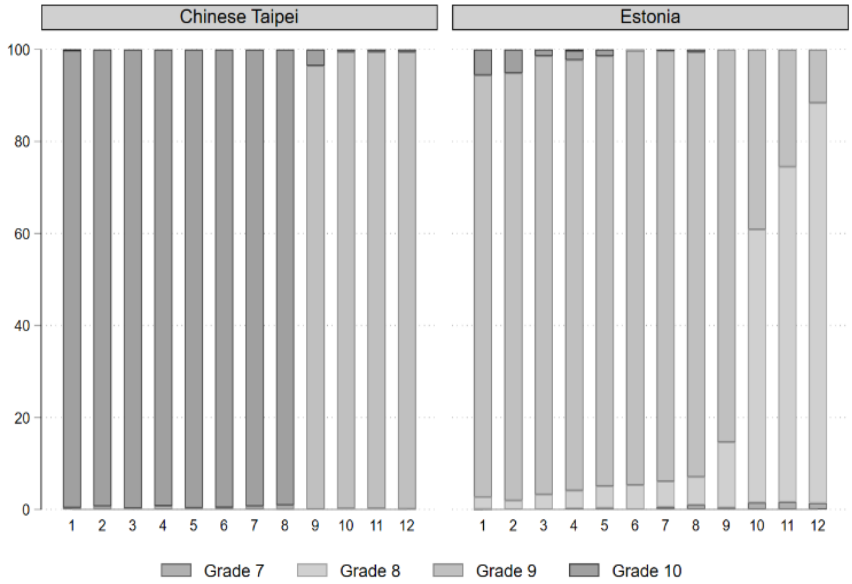
in particular months of the year (e.g. if those born in the two months following the cut-off date may apply for early enrolment under certain conditions). Non-linearities that have their origin in administrative rules are addressed by excluding students born in certain months from the analysis. In particular, for Costa Rica and Baku (Azerbaijan), students born in the cut-off month are excluded because the cut-off date (the 15th of that month) does not allow for the computation, with certainty, of an expected grade for these students. For Germany and Switzerland, all months of birth between the earliest local cut-off date and the latest local cut-off date are excluded from the analysis (see Footnote 7). For the Czech Republic and the Slovak Republic, all students born from June through August are excluded from the analysis. In the latter two countries, children who turn six before 31 August must pass an interview with the school to determine whether they are ready for primary school attendance, and children born in June, July and August (i.e. the youngest students) are more often deemed not ready compared to students born in the remaining months.

3.2. Descriptive statistics

In all countries and economies included in this study, the PISA test was conducted at a time when the PISA-eligible population was expected to be found in two contiguous grades rather than in a single grade. For students in the PISA cohort of these countries and economies, the theoretical school starting age does not increase as a function of the age at the time of the PISA test but shows a sharp discontinuity. Indeed, some of the youngest students who sat the PISA test in 2018 or 2015 (those who were not much older than 15 years and 3 months) were among the oldest students in their first-grade class; and some of the oldest students who sat the PISA test were among the youngest students in their first-grade class. When the PISA test was conducted, the youngest students in the PISA cohort were expected to be in a lower grade (e.g. Grade 9) than the oldest students (Grade 10).

Figure 3.1 illustrates the consequences of the misalignment between the cut-off dates for first-grade enrolment and the cut-off dates defining the PISA cohort with Estonia and Chinese Taipei, two jurisdictions that conducted the PISA test (in both years) around the month of April, so that the PISA-eligible cohorts coincided with students born in 1999 and 2002 (January through December). In Estonia, the cut-off date for enrolment in first grade is 1 October; in Chinese Taipei, it is 1 September. In both Estonia and Chinese Taipei, a clear discontinuity in the grade attended by students assessed in PISA is visible around the cut-off date for school entry. In detail, school-entry regulations almost perfectly predict the grade of PISA students about ten years later in Chinese Taipei and are strongly (but not perfectly) related to students’ grades in Estonia.

Figure 3.1. Grade distribution per month of birth in Estonia and Chinese Taipei



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

Table 3.2 presents similar information for all countries and economies analysed in the present study. In all countries and economies, the youngest and the oldest student are *almost* one year of age and one full grade level apart but are close in terms of their expected age at school entry. For most countries, there is a sharp discontinuity, around the age which corresponds to children born just after the cut-off date for first-grade enrolment, in the number of grade levels completed at the time of the PISA test.

Table 3.2. Average grade level by age in selected countries and economies

Country/economy	15y3m	15y4m	15y5m	15y6m	15y7m	15y8m	15y9m	15y10m	15y11m	16y	16y1m	16y2m
Albania	9.1	9.1	9.2	9.5	**9.7**	9.7	9.8	9.9	9.9	9.9	9.9	9.9
Baku (Azerbaijan)	9.0	9.2	9.4	9.6	**9.7**	9.7	9.7	9.8	9.8	9.8	9.8	9.9
Belarus	8.9	9.0	9.0	9.1	**9.6**	9.7	9.8	9.8	9.9	9.9	9.9	9.9
B-S-J-Z (China)	9.1	9.1	9.2	9.3	**9.8**	9.8	9.8	9.9	9.9	9.9	9.9	10.0
Costa Rica	8.7	8.7	8.8	9.1	**9.3**	9.4	9.4	9.5	9.5	9.4	9.4	9.5
Croatia	9.0	9.0	9.0	9.0	9.0	9.0	9.0	9.1	9.1	**9.7**	9.8	9.9
Czech Republic	8.9	8.9	8.9	9.0	9.3	9.5	**9.6**	9.6	9.7	9.8	9.8	9.8
Estonia	8.1	8.2	8.4	**8.8**	8.9	8.9	9.0	9.0	9.0	9.0	9.0	9.0
Finland	8.0	**8.8**	8.9	8.9	9.0	9.0	9.0	9.0	9.0	9.0	9.0	9.0
Germany	8.9	8.9	9.0	**9.1**	9.2	9.4	9.6	9.7	9.7	9.7	9.7	9.7
Hong Kong (China)	8.9	**9.5**	9.5	9.6	9.6	9.7	9.7	9.7	9.7	9.7	9.8	9.8
Hungary	8.7	8.8	8.8	8.9	8.9	8.9	9.0	**9.2**	9.2	9.4	9.4	9.4
Korea	9.1	9.2	**10.0**	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0
Luxembourg	8.9	8.9	8.9	9.0	**9.5**	9.5	9.5	9.5	9.6	9.6	9.6	9.6
Serbia	9.0	9.0	9.0	9.0	9.0	9.0	9.0	9.0	9.0	9.1	**9.6**	9.7
Slovak Republic	8.9	8.9	8.9	9.0	**9.4**	9.6	9.7	9.7	9.8	9.8	9.8	9.8
Switzerland	8.9	8.9	8.9	9.0	9.0	**9.0**	9.1	9.2	9.4	9.5	9.6	9.6
Chinese Taipei	9.0	9.0	9.0	9.0	**10.0**	10.0	10.0	10.0	10.0	10.0	10.0	10.0

Notes: "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. Cells with a grey background indicate months of births that are excluded from the analysis. Bold font indicates the age of the youngest PISA students who are expected to be in a higher grade, according to school-entry regulations (see Table 3.1).
Source: OECD PISA 2015 and 2018 datasets, https://www.oecd.org/pisa/data/ (accessed on 17 May 2021).

Table 3.3 presents the mean performance of each country and economy, along with standard deviations of test scores. As can be seen, the PISA participants included in this study span a wide range of performance (mean performance in reading, for example, ranges from below 400 score points in Baku [Azerbaijan] to over 550 points in B-S-J-Z [China]). Across all countries and economies, the standard deviation of test scores is comprised between 73 and 108 score points; this can be used to convert score differences into an effect-size metric.

Table 3.3. Mean score and variation in performance in PISA in selected countries and economies

Country/economy	PISA 2015						PISA 2018					
	Reading		Mathematics		Science		Reading		Mathematics		Science	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Albania	405	97	413	86	427	78	405	80	437	83	417	74
Baku (Azerbaijan)	m	m	m	m	m	m	389	74	420	89	398	74
Belarus	m	m	m	m	m	m	474	89	472	93	471	85
B-S-J-Z (China)	m	m	m	m	m	m	555	87	591	80	590	83
Costa Rica	427	79	400	68	420	70	426	81	402	75	416	73
Croatia	487	91	464	88	475	89	479	89	464	87	472	90
Czech Republic	487	100	492	91	493	95	490	97	499	93	497	94
Estonia	519	87	520	80	534	89	523	93	523	82	530	88
Finland	526	94	511	82	531	96	520	100	507	82	522	96
Germany	509	100	506	89	509	99	498	106	500	95	503	103
Hong Kong (China)	527	86	548	90	523	81	524	99	551	94	517	86
Hungary	470	97	477	94	477	96	476	98	481	91	481	94
Korea	517	97	524	100	516	95	514	102	526	100	519	98
Luxembourg	481	107	486	94	483	100	470	108	483	98	477	98
Serbia	m	m	m	m	m	m	439	96	448	97	440	92
Slovak Republic	453	104	475	95	461	99	458	100	486	100	464	96
Switzerland	492	98	521	96	506	100	484	103	515	94	495	97
Chinese Taipei	497	93	542	103	532	100	503	102	531	100	516	99

Notes: SD: Standard deviation. m: missing (the country/economy did not participate in PISA 2015).
Source: OECD PISA 2015 and 2018 datasets, https://www.oecd.org/pisa/data/ (accessed on 17 May 2021).

4. Identification strategy

The learning progression observed over a year for each student is the joint effect of age (maturity) and of the learning experiences accumulated over this period, among which school instruction plays an important role. Individual skills measured in tests reflect the accumulation of these yearly progressions over time, but inferring the typical, average progressions from comparisons of individuals who differ in terms of age and years of schooling presents many challenges. First, in a cross-sectional, single-cohort study such as the Programme for International Student Assessment (PISA), there is only limited variation in test-takers' age. But more importantly, the variation in their grade levels (i.e. in their years of schooling) is often related to unobserved factors that may also exert a direct influence on test scores. As a result, the identification of grade-and-age effects is challenging and not always possible.

For countries where there is no single expected grade in the PISA cohort, the difference in PISA scores across these two grades can be identified under the assumption that all age effects on test scores are locally linear.

There are two kinds of age effects on test scores: age-at-the-test and school-starting-age effects. Age-at-the-test effects refer to the effect of sitting a particular test at an older age; for students who are tested at the same point in their school career, age differences at the time of the test are related to greater maturity and, possibly, to a greater number of out-of-school learning experiences. School-starting-age effects refer to the effect of starting school at an older age; differences in school starting age may affect test scores even later in the school career because initial differences in maturity may have lasting effects on children's school career and personality (Givord, 2020[24]). In a cross-sectional study, such as PISA, it is usually not possible to distinguish the two age effects, as well as length-of-schooling effects, separately: indeed, age at school start, the number of grade levels completed, and the current age of the student are linked by a simple, additive relationship.[11]

Assuming that both age effects are linear, if all children comply perfectly with age-at-entry regulations and progress regularly through primary and secondary schooling, then the learning gain over one year of schooling (grade gain) can be measured in a simple regression analysis of performance on grade ($grade_{is}$, an indicator variable equal to 1 if the student is in the upper of the two grades, and equal to 0 otherwise), which also accounts for students' school starting age (in months), ssa_{is}. The coefficient on the "grade" indicator corresponds to the combined effect of being in a higher grade (having completed one additional year of schooling) and of being one year older.

More specifically, one can estimate the parameters of the following linear regression equation:

Equation 4.1

$$y_{is} = \alpha + \beta * ssa_{is} + \gamma * grade_{is} + \delta' * x_{is} + \epsilon_{is}$$

However, in practice, many students are held back during their school careers, skip grades, or do not perfectly comply with regulations about the age at entry into school. All of these decisions, taken by students' families, teachers, or by doctors and psychologists, may themselves have been based on factors that predict performance at age 15 (such as students' health or cognitive maturity), and that cannot be observed in PISA data (they are not included in the x_{is} vector of control variables). The observed grade ($grade_{is}$) is therefore endogenous, and the apparent effects of grades may be confounded by these other factors. In addition, the exact school starting age of each student is not directly observed in PISA.[12]

The grade gain may still be identified in an instrumental-variables regression, which exploits the variation in grades that stems solely from regulations. More specifically, one can use a student's theoretical grade (which is assumed to be exogenous, independent of those other factors that may predict performance) to instrument the actual observed grade (which is endogenous). Similarly, one can use the theoretical school-starting age in place of the actual school-starting age (which is unknown). The key assumptions in this approach are that the theoretical grade, a function of the month of birth, is strongly related to the actual observed grade and that it influences performance in PISA only through its influence on the observed grade.

In practice, there may be other reasons why a student's month of birth influences his or her performance – besides differences in school-starting age, maturity and in the actual length

[11] In PISA, the testing date is not exactly identical for all students. Therefore, in theory, variation in age-at-the-test for students born in the same month may help distinguish school-starting-age effects and age-at-the-test effects. In practice, however, this variation is very limited (at most, eight weeks), and distinguishing the two age effects is refrained from. Instead, age at the test is treated as equal to the sum of the (theoretical) school-starting age and the (theoretical) number of completed grades for each student.

[12] The PISA questionnaire asks students to report the age at which they started primary school. However, such retrospective questions are subject to reporting errors, referred to as "telescoping" errors in survey research (Tourangeau and Bradburn, 2010, pp. 330-332[50]).

of schooling. In particular, some months of birth may also be associated with particular family characteristics that are not included in the x_{is} vector of control variables, but are consequential for students' learning and for their academic progression. For example, certain parents may try to manipulate their children's birth date precisely because of the existence of strict school-entry regulations; this threat to the identification of the grade gain is addressed in robustness analyses that exclude the months immediately preceding and following the cut-off date from the estimation samples.

The empirical specification of Equation 4.1 includes a year-specific constant and the following control variables x_{is}: gender, immigrant background and the economic, social and cultural status (in quarters, to account for the different scales of the variable in 2015 and 2018).

4.1. Limitations and robustness checks

The identification of age-and-grade effects based on the difference in the cut-off dates that determine eligibility for first-grade enrolment and for PISA participation critically relies on two assumptions: that a student's birthday influences test scores only through age effects (age-at-test and school-starting age), and that age-at-test and school-starting-age effects are linear.

The first assumption – an exogeneity assumption required for the instrumental-variable estimation – is violated if family characteristics differ in unobserved but impactful ways across birth dates, for example, if a particular type of parent chooses to schedule the delivery of their baby either just before or just after the cut-off date for first-grade enrolment (or just before or after 1 January, which in most countries is not only the beginning of a new calendar year but also the cut-off for PISA eligibility).

The second assumption (linearity) means, for instance, that compared to students who were born in the fifth month of the twelve-month period defining a school cohort, the advantage of being born in the first month (i.e. being four months older) is of the same magnitude as the disadvantage of being born in the ninth month (i.e. being four months younger). It is plausible to assume that age-at-test effects, which have a physiological origin (maturity), are approximately linear around the age of 15.[13] However, school-starting-age effects do not only reflect initial differences in maturity but also the consequences of such initial differences on students' careers, which are not necessarily proportional to the initial difference in age. Non-linearities may result from peer effects – for example, if the effect of having high-ability peers is mostly limited to students at the high end of the ability distribution; see (Sacerdote, 2011[25])); or from institutional selection mechanisms – for example, in systems with early tracking, the probability of being selected into the most prestigious school track may not increase linearly with age. They may also result from the existence of exceptions to the cut-off dates for primary school entry that target only students born in particular months (for example, if students born in the month preceding

[13] Some scholars, however, have shown that physiological differences between students born in different months of the year do not necessarily follow a linear, or even monotonic pattern. Non-linearities in age effects may reflect differences in maternal characteristics by month of birth (Buckles and Hungerman, 2013[45]); or differences in in-utero conditions related to the month of birth, e.g. for children whose mothers were pregnant, and fasting, during the month of Ramadan (Majid, 2015[46]; Almond and Mazumder, 2011[47]).

the cut-off date can anticipate enrolment when places are available, or when a qualified professional certifies their maturity for school).[14]

As a robustness check, where possible, a set of estimates is provided based on restricted samples that exclude students born in the months preceding and following the cut-off date for first-grade enrolment (e.g. excluding students born in August and in September, in countries where the cut-off date is 1 September). These restricted-sample estimates address the most likely violations of the exogeneity assumption and of the (strict) linearity assumption, which are both expected to occur around the cut-off for school entry. The smaller samples, however, are expected to result in inflated confidence intervals, reflecting greater statistical uncertainty, in situations where these assumptions are not violated.

For some countries and economies, it is also possible to relax the (strict) linearity assumption and to let the slope of the linear relationship between age and performance differ for the eldest and youngest students in PISA (who are, respectively, the youngest and eldest in their school-entry cohort). This is possible, in particular, when there are two or more months in the PISA cohort on either side of the first-grade-entry cut-off; the regression function that is estimated corresponds, then, to that of a typical regression-discontinuity design, where the "running variable" is the school-starting age and a discontinuity in students' age and grade is observed in correspondence with the school-starting age of the eldest students eligible to take the PISA test.

In detail, let $y_{is}(0)$ be the test score of a student in the lower of two grades (and ages), and $y_{is}(1)$ be the test score of the same student in the higher of two grades; only one of these is observed in PISA. The quantity of interest is $y_{is}(1) - y_{is}(0)$; this quantity can be identified, under certain assumptions, at the cut-off date that defines eligibility for participation in PISA. In particular, let M indicate the (expected) school starting age of students born on this cut-off date (e.g. 1 January, in countries where the PISA cohort coincides with all students born in a particular year). If all students complied with school-entry regulations, one would observe $y_{is} = y_{is}(1)$ if $ssa_{is} > M$ and $y_{is} = y_{is}(0)$ if $ssa_{is} < M$. Further assume that age effects are linear within each grade level but not necessarily parallel across grade levels:

Equation 4.2

$$\begin{cases} y_{is}(0) = \alpha_0 + \beta_0 * (ssa_{is} - M) + \epsilon_{is} \\ y_{is}(1) = (\alpha_0 + \alpha_1) + (\beta_0 + \beta_1) * (ssa_{is} - M) + \epsilon_{is} \end{cases}$$

The average grade gain for students whose birth date is such that $ssa_{is} = M$, $E(y_{is}(1) - y_{is}(0) | ssa_{is} = M)$is then equal to:

[14] For the Czech Republic, Germany, the Slovak Republic and Switzerland, the variation or flexibility of administrative rules also results in violations of this linearity assumption. This is addressed by excluding students born in certain months from the analysis. The linearity assumption must therefore hold only for the remaining months-of-birth.

Equation 4.3

$$\lim_{ssa_{is}\uparrow M} \alpha_0 + \beta_0 * (ssa_{is} - M) - \lim_{ssa_{is}\downarrow M} (\alpha_0 + \alpha_1) + (\beta_0 + \beta_1) * (ssa_{is} - M) = \alpha_1$$

This quantity can be estimated by fitting the following regression function to the observed test scores:

Equation 4.4

$$y_{is} = \alpha_0 + \beta_0 * (ssa_{is} - M) + \alpha_1 \mathbb{I}_{ssa_{is}>M} + \beta_1 * (ssa_{is} - M) * \mathbb{I}_{ssa_{is}>M} + \epsilon_{is}$$

Note that $\mathbb{I}_{ssa_{is}>M}$ (an indicator function) is equivalent, up to a constant, to the expected grade, given school-entry regulations. In the presence of imperfect compliance with school-entry regulations and of grade repetition, a "fuzzy regression-discontinuity" estimator can be obtained by using $\mathbb{I}_{ssa_{is}>M}$ as an instrumental variable for $grade_{is}$ in the following equation:

Equation 4.5

$$y_{is} = \alpha_0 + \beta_0 * (ssa_{is} - M) + \alpha_1 * grade_{is} + \beta_1 * (ssa_{is} - M) * \mathbb{I}_{ssa_{is}>M} + \epsilon_{is}$$

Finally, a third check consists in testing the absence of differences between the eldest and the youngest students in the PISA cohort, other than age-and-grade differences. The exogeneity assumption implies that all unobserved determinants of performance that are not included as control variables in the regression are as good as randomly distributed across months of birth, and in particular around the discontinuity of interest. If key variables that are predictive of students' performance – such as students' socio-economic status, gender and immigrant background – do not vary significantly between the eldest and the youngest students' in PISA, this would reinforce one's confidence that other potential confounders, which are unobserved and can therefore not be included as control variables, also do not vary. The absence of selection effects is tested for by estimating a version of Equation 4.1 where the socio-demographic control variables replace test scores as the dependent variable (no controls are included other than age and grade, instrumented by the expected grade given school-entry regulations).

Finally, as usual with instrumental-variables estimators, the coefficients on the grade variable (γ in Equation 4.1 or α_1 in Equation 4.5) must be interpreted as a "local average treatment effect" and cannot be generalised to the full sample or to the population from which the sample is drawn without additional assumptions. In particular, in situations where families do not perfectly comply with school-entry regulations, or where students may be held back in their progress through the school grades, the instrumental-variable estimate must be interpreted as the average effect on the population whose grade level in PISA varies as a function of their birthday. This population excludes all students who would have been enrolled in a higher grade, irrespective of their birthday (e.g. as a result of grade skipping or early enrolment); and all students who would have been enrolled in a lower grade, irrespective of their birthday (e.g. as a result of grade retention or delayed enrolment). The first-stage coefficient, which corresponds to the change in the proportion of PISA students in the upper grade around the discontinuity in the school-starting-age-to-grade mapping, indicates the size of the population of "compliers"; the larger this coefficient, the more "representative" the estimated grade gains.[15]

5. Results

5.1. The grade gain in mathematics, reading and science

Table 5.1 shows the average grade gain estimated using the expected variation in gains based solely on administrative school-entry regulations as an "instrumental variable".

Table 5.1. Grade-and-age effects in 18 countries and economies

Country/economy and years	Grade-and-age effect[1]						Number of observations	First stage: Effect of the expected grade on the actual grade[2]	
	Mathematics		Reading		Science				
Albania (2018)	1.9	(6.6)	9.2	(5.8)	10.6	(5.9)	6 202	0.82	(0.02)
Baku (Azerbaijan)[3] (2018)	22.1	(8.4)	12.5	(5.7)	12.6	(6.4)	5 805	0.67	(0.02)
Belarus (2018)	16.7	(6.0)	10.0	(4.6)	10.9	(4.6)	5 631	1.03	(0.01)
B-S-J-Z (China) (2018)	12.9	(5.9)	17.0	(5.6)	17.2	(6.4)	11 950	0.82	(0.02)
Costa Rica[3] (2015, 2018)	18.7	(5.2)	27.5	(5.1)	25.3	(5.5)	12 397	0.61	(0.02)
Croatia (2015, 2018)	19.0	(3.5)	17.3	(4.1)	14.4	(3.7)	12 093	0.84	(0.01)
Czech Republic[3] (2015, 2018)	10.7	(5.6)	16.4	(5.4)	13.2	(5.3)	10 004	0.92	(0.01)
Estonia (2015, 2018)	37.9	(4.3)	27.2	(4.6)	29.8	(4.5)	10 610	0.78	(0.01)
Finland (2015, 2018)	16.0	(3.8)	17.8	(3.9)	20.2	(4.4)	11 285	0.97	(0.01)
Germany[3] (2015, 2018)	34.2	(7.2)	25.6	(7.5)	33.1	(7.5)	6 621	0.68	(0.02)
Hong Kong (China) (2015, 2018)	27.5	(5.3)	20.4	(5.7)	23.5	(4.7)	10 920	0.85	(0.02)
Hungary (2015, 2018)	15.5	(7.4)	17.1	(7.8)	8.9	(8.0)	10 604	0.49	(0.02)
Korea (2018)	18.8	(8.3)	8.7	(7.2)	12.3	(8.4)	6 568	0.88	(0.01)
Luxembourg (2015, 2018)	31.3	(6.9)	30.1	(7.0)	23.2	(6.6)	10 093	0.56	(0.01)
Serbia (2018)	9.5	(7.1)	18.1	(7.1)	20.3	(7.2)	6 378	0.67	(0.02)
Slovak Republic[3] (2015, 2018)	19.1	(5.3)	31.0	(5.5)	24.9	(4.4)	8 700	0.88	(0.02)
Switzerland[3] (2015, 2018)	35.6	(9.5)	35.8	(9.9)	30.3	(10.1)	7 583	0.59	(0.02)
Chinese Taipei (2015, 2018)	10.9	(4.8)	7.6	(4.0)	8.4	(4.5)	14 736	0.99	(0.00)

Notes: Each row corresponds to a separate regression. All regressions include, in addition to grade and school-starting-age variables, controls for students' socio-economic status (three dummies, for quarters), gender and immigrant background. For countries/economies whose estimates are based on the pooled 2015 and 2018 data, a year dummy is also included. 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.

[15] Even in a situation of perfect compliance, the estimated effects are "local" in another sense: they refer, strictly speaking, only to students born just before or just after the cut-off date defining eligibility for PISA. It is, however, reasonable to assume that the grade gain at age 15 is unrelated to a students' birthday and school starting age, and can thus be generalised to students born in the remaining months who share similar characteristics. It can also be noted that the "local" interpretation of the instrumental variable coefficient requires an assumption of "monotonicity", i.e. the assumption that (around the cut-off date), a later birthday either increases the probability of being in a lower grade or leaves this probability unchanged (e.g. at 0% or 100%), but in no case decreases it (Barua and Lang, 2016[49]). There is no reason to believe that there are "defiers" in this case, i.e. students who would be in a lower grade, had they been born earlier; or in a higher grade, had they been born later.

1. Grade-and-age effects for each subject are estimated using separate instrumental-variable regressions. They correspond to the coefficient on the grade variable (instrumented by the theoretical grade). See Equation 4.1 for details.
2. The first-stage regression is common to all subjects; it also includes school-starting age and other exogenous variables (control variables), whose coefficients are not reported.
3. The analysis does not include students born in certain months. See Section 4 for details.
Source: OECD PISA 2015 and 2018 datasets, https://www.oecd.org/pisa/data/ (accessed on 17 May 2021).

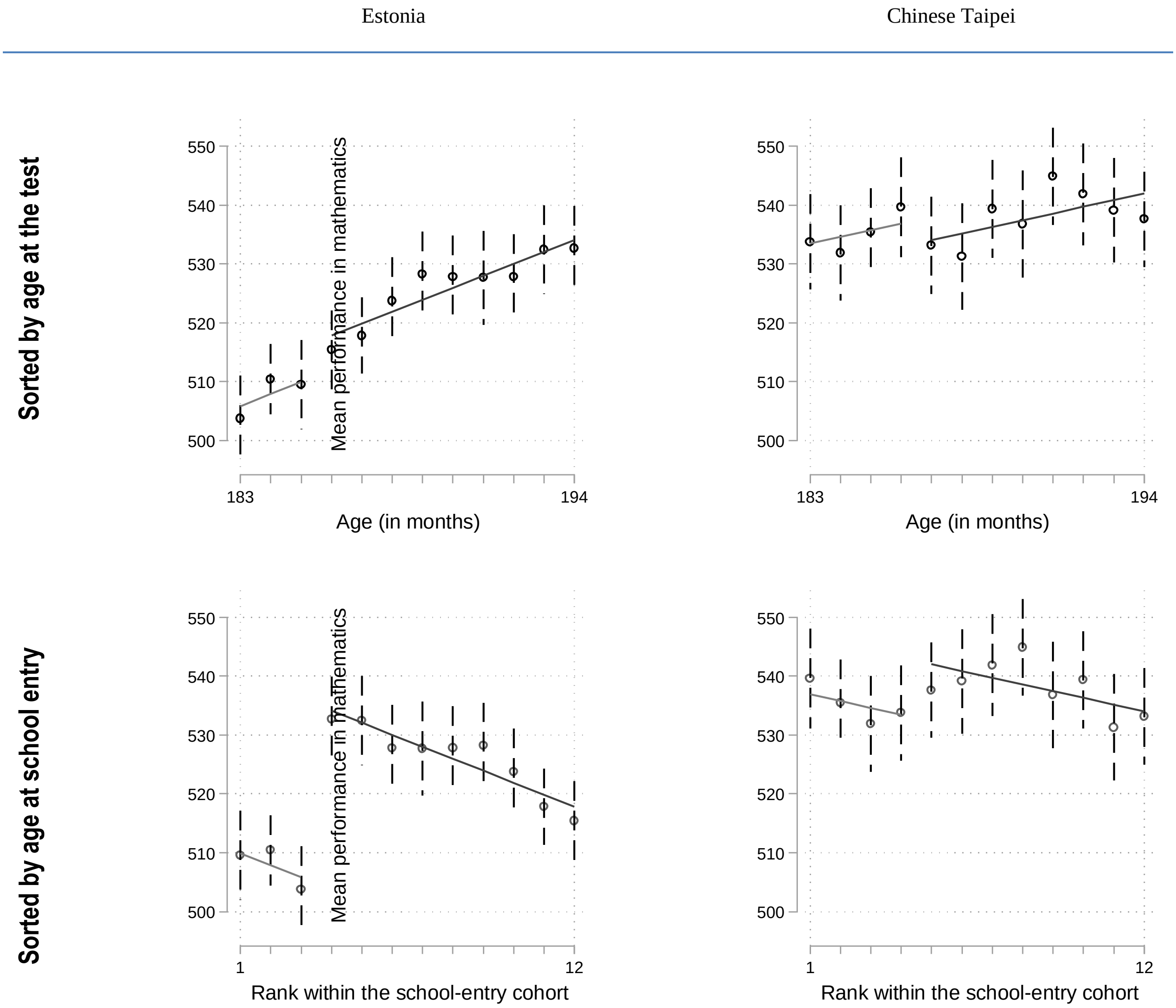
Grade-and-age effects for mathematics reported in Table 5.1 range from below 10 points (and not statistically significantly different from 0) in Albania and Serbia to more than 30 points in Estonia, Germany, Luxembourg and Switzerland.

In all three subject domains, grade-effect estimates are, for a majority of countries, comprised between 10 and 25 score points. Given the relatively large standard errors that accompany some of these estimates, only large between-country differences tend to be statistically significant.

A simple graphical representation shows the intuition behind the instrumental-variable estimator presented in Table 5.1 and shows in what sense this estimator can be interpreted as a regression-discontinuity estimator.

Figure 5.1 shows the average performance of students in Estonia and Chinese Taipei by month of birth. Given that month of birth is directly related to the instrument used (the expected grade at the time of the PISA test), this figure represents a reduced-form version of the instrumental-variable regression.

Figure 5.1. Mean performance in mathematics by month of birth in Estonia and Chinese Taipei



Notes: The same data points are shown in the top and bottom charts, which differ in the sorting order. Vertical dashed lines indicate the 95% confidence interval around the mean. Regression lines show the expected performance based on students' age and theoretical grade level. The two lines have the same slope; each line refers to a group of students expected to be in a particular grade, given local school-entry regulations.
Source: OECD PISA 2015 and 2018 datasets, https://www.oecd.org/pisa/data/ (accessed on 17 May 2021).

For each country and economy, the same data are presented in two distinct ways.

In the first plot, birth months are sorted by students' age at the PISA test; the left-most dot corresponds to the mean performance of the youngest students in the PISA cohort (students born in December, in both cases). In the second plot, just below, birth months are sorted by their rank in the school-entry cohort. The left-most dot corresponds to the mean performance of the oldest students within their school-entry cohort, given local regulations about school entry (students born in October, in Estonia; and in September, in Chinese Taipei). The two segments of the regression line have the same slope (corresponding to a reduced-form estimate of age effects, which are assumed to be linear).

In the top plot, the "jump" observed at the cut-off date for school entry corresponds to the reduced-form coefficient on the grade variable: this is the effect of being in a higher grade, plus the effect of being the youngest (as opposed to the eldest) within one's school-entry cohort. Because the latter effect (which is usually negative) may be as strong as the penalty associated with being in a higher grade (absent differences in age), the net result may be positive, as in Estonia, or negative, as in Chinese Taipei.

The second plot, which displays the same data and regression lines, but ranked by the students' age at school entry, visually highlights the source of grade-and-age effects reported in Table 5.1. This time, the discontinuity corresponds to students who are (almost) one year older (born in January, as opposed to December, of the same year) and one grade further advanced in their schooling – but have almost the same school-starting age. This reduced-form effect corresponds to almost 30 score points in Estonia and about 10 score points in Chinese Taipei and is close, given the strong compliance with school-entry regulations observed in both systems (see Figure 3.1), to the instrumental-variable estimate reported in Table 5.1.

5.2. Robustness checks

Table 5.2 presents an alternative set of estimates based on the same specification as in Table 5.1, but where the estimation sample excludes the eldest and youngest students in the (theoretical) school-entry cohort (i.e. the right-most and left-most month of birth, in the bottom charts of Figure 5.1). This is similar to using a reduced bandwidth in a regression-discontinuity design. Table 5.2 constitutes a robustness check in the sense that it requires less stringent assumptions about the linearity of school-starting-age effects and the exogeneity of birth dates; in particular, the linearity assumptions can be relaxed for students born less than one month before or after the cut-off date defining a school-entry cohort.

The point estimates in Table 5.2 are, for the large majority of countries and economies, very close to those presented in Table 5.1. Differences in excess of 2.5 score points are found for Baku (Azerbaijan), Belarus, Germany, Korea and Serbia only and are often limited to one subject; among these countries, they are smaller than the estimates in Table 5.1 in Germany and Korea only.

Table 5.3 presents estimates that relax the strict linearity assumption embedded in Equation 4.1 and allow for the relation between school-starting age and test scores (and thus also between age and test scores) to vary within each grade (see Equation 4.5). This corresponds to a more typical regression-discontinuity estimator, where separate regression functions approximate the relationship between the running variable (school-starting age, in this case) and the outcome on either side of the discontinuity. In most countries and economies, the more flexible functional form results in wider confidence intervals (particularly in countries where the piece-wise linear age effects must be estimated on a small number of birth months, such as Korea and Serbia), but in similar point estimates to those reported in Table 5.3. It must be noted, however, that it is not possible to use this more flexible estimator in Finland and Hong Kong (China) because one side of the discontinuity includes only students born in a single month; and that this more flexible function results in a weak first-stage relationship in Switzerland, whose results are therefore not reported.

Table 5.2. Robustness check: Grade-and-age effects estimated on restricted samples

Excluding students born in the months preceding and following the cut-off date from the estimation sample

Country/economy and years	Grade-and-age effect[1]						Number of observations[2]	First stage: Effect of the expected grade on the actual grade[3]	
	Mathematics		Reading		Science				
Albania (2018)	3.1	(7.3)	9.9	(6.4)	12.9	(6.5)	5 226	0.81	(0.02)
Baku (Azerbaijan)[4] (2018)	21.1	(8.2)	13.2	(5.5)	16.6	(6.6)	4 613	0.72	(0.02)
Belarus (2018)	21.0	(6.3)	11.3	(5.0)	12.4	(4.8)	4 640	1.01	(0.01)
B-S-J-Z (China) (2018)	13.8	(6.2)	19.5	(6.1)	17.8	(6.9)	9 927	0.82	(0.02)
Costa Rica[4] (2015, 2018)	16.8	(5.5)	28.9	(5.7)	25.9	(6.1)	9 962	0.61	(0.02)
Croatia (2015, 2018)	18.5	(4.0)	18.3	(4.4)	14.4	(4.0)	10 165	0.85	(0.01)
Czech Republic[4] (2015, 2018)	10.4	(5.5)	16.3	(5.4)	13.1	(5.2)	8 979	0.92	(0.01)
Estonia (2015, 2018)	35.9	(4.5)	26.4	(4.5)	30.5	(4.5)	8 834	0.81	(0.01)
Finland (2015, 2018)	c	c	c	c	c	c	c	c	c
Germany[4] (2015, 2018)	28.9	(7.8)	22.8	(8.5)	28.6	(9.3)	5 000	0.67	(0.02)
Hong Kong (China) (2015, 2018)	c	c	c	c	c	c	c	c	c
Hungary (2015, 2018)	14.7	(8.9)	15.9	(9.1)	7.1	(9.1)	8 893	0.48	(0.02)
Korea (2018)	14.7	(11.3)	6.4	(9.1)	11.5	(10.8)	5 544	0.91	(0.01)
Luxembourg (2015, 2018)	30.3	(7.7)	29.1	(7.6)	21.8	(7.3)	8 431	0.55	(0.02)
Serbia (2018)	15.1	(7.9)	22.2	(8.1)	24.9	(8.5)	5 384	0.72	(0.02)
Slovak Republic[4] (2015, 2018)	19.1	(5.5)	30.1	(5.6)	24.0	(4.4)	7 707	0.87	(0.02)
Switzerland[4] (2015, 2018)	37.9	(10.4)	37.1	(10.8)	32.4	(10.4)	5 577	0.59	(0.02)
Chinese Taipei (2015, 2018)	10.7	(5.0)	7.2	(4.4)	7.8	(4.7)	12 151	0.99	(0.00)

Notes: Each row corresponds to a separate regression. All regressions include, in addition to grade and school-starting-age variables, controls for students' socio-economic status (three dummies, for quarters), gender and immigrant background. For countries/economies whose estimates are based on the pooled 2015 and 2018 data, a year dummy is also included. 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.

1. Grade-and-age effects for each subject are estimated using separate instrumental-variable regressions. They correspond to the coefficient on the grade variable (instrumented by the theoretical grade). See Equation 4.1 for details.
2. All regressions are estimated on samples that include, at most, students born in 10 out of 12 months.
3. The first-stage regression is common to all subjects; it also includes school-starting age and other exogenous variables (control variables), whose coefficients are not reported.
4. The analysis includes fewer than 10 months of birth. See Section 4 for details.

c: There are no observations to estimate Equation 4.1 after excluding the months adjacent to the cut-off date from the sample (all remaining students are expected to be in the same grade).

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

Table 5.3. Robustness check: A regression-discontinuity estimator of grade-and-age effects

Assuming different age effects for the eldest and youngest students in PISA

Country/economy and years	Grade-and-age effect[1]						Number of observations	First stage: Effect of the expected grade on the actual grade[2]	
	Mathematics		Reading		Science				
Albania (2018)	0.0	(6.4)	6.5	(5.7)	6.9	(6.3)	6 202	0.95	(0.02)
Baku (Azerbaijan)[3] (2018)	23.3	(7.8)	15.3	(5.6)	17.0	(6.8)	5 805	0.87	(0.02)
Belarus (2018)	24.8	(6.9)	13.5	(5.1)	13.7	(5.3)	5 631	1.04	(0.01)
B-S-J-Z (China) (2018)	8.6	(6.4)	14.4	(5.8)	12.6	(6.7)	11 950	0.88	(0.02)
Costa Rica[3] (2015, 2018)	16.3	(6.2)	30.6	(5.8)	23.8	(6.2)	12 397	0.64	(0.02)
Croatia (2015, 2018)	18.3	(4.8)	20.5	(5.0)	14.2	(4.7)	12 093	0.96	(0.02)
Czech Republic[3] (2015, 2018)	11.2	(5.8)	17.1	(5.6)	13.5	(5.7)	10 004	0.90	(0.01)
Estonia (2015, 2018)	32.5	(5.0)	23.7	(4.9)	27.1	(4.2)	10 610	0.95	(0.01)
Finland (2015, 2018)	c	c	c	c	c	c	c	c	c
Germany[3] (2015, 2018)	32.2	(8.3)	24.8	(8.2)	30.9	(8.4)	6 621	0.70	(0.02)
Hong Kong (China) (2015, 2018)	c	c	c	c	c	c	c	c	c
Hungary (2015, 2018)	19.2	(7.0)	18.4	(7.6)	11.0	(7.6)	10 604	0.55	(0.02)
Korea (2018)	17.7	(15.0)	10.2	(12.2)	17.9	(14.6)	6 568	0.95	(0.02)
Luxembourg (2015, 2018)	30.7	(7.9)	31.8	(8.3)	24.0	(7.6)	10 093	0.56	(0.02)
Serbia (2018)	18.0	(10.3)	24.4	(10.0)	28.3	(10.6)	6 378	0.83	(0.04)
Slovak Republic[3] (2015, 2018)	18.9	(5.4)	32.2	(5.7)	25.3	(4.7)	8 700	0.88	(0.02)
Switzerland[3] (2015, 2018)	32.1	(10.7)	33.3	(11.0)	28.2	(10.6)	7 573	0.61	(0.03)
Chinese Taipei (2015, 2018)	12.5	(5.0)	7.7	(4.2)	6.7	(4.5)	14 736	1.00	(0.00)

Notes: Each row corresponds to a separate regression. All regressions include, in addition to grade and school-starting-age variables, controls for students' socio-economic status (three dummies, for quarters), gender and immigrant background; for countries/economies whose estimates are based on the pooled 2015 and 2018 data, a year dummy is also included. In this table, the school-starting-age variable is also interacted with the expected grade dummy to allow for different slopes of the regression function on either side of the discontinuity in the school-starting-age/age-in-PISA mapping; see Equation 4.5 for details. 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.
1. Grade-and-age effects for each subject are estimated using separate instrumental-variable regressions. They correspond to the coefficient on the grade variable (instrumented by the theoretical grade) in Equation 4.5.
2. The first-stage regression is common to all subjects; it also includes, in addition to the expected grade, two school-starting-age variables (school-starting-age and its interaction with the expected grade) and other exogenous variables (control variables), whose coefficients are not reported.
3. The analysis includes fewer than 12 months of birth. See Section 4 for details.
c: There are insufficient observations to estimate separate school-starting-age effects for the eldest and youngest students in PISA (all students expected to be in a higher grade have the same age, in months).
Source: OECD PISA 2015 and 2018 datasets, https://www.oecd.org/pisa/data/ (accessed on 17 May 2021).

Finally, Table 5.4 shows that results reported in Table 5.1 are unlikely to be driven by selection effects and are most likely to reflect only the combined effect of age and length of schooling. Indeed, all three control variables included in the regression show limited, if any, variation between students who are older, and in a higher grade, and students who are younger, and in a lower grade. Such pseudo-grade-and-age effects could result from either seasonal birth patterns (or birth-date manipulation), whereby, for example, the families of students born in January differ systematically from the families of students born in December; or from sample-selection issues, due, for example, to students dropping out of school on their 16th birthday.

Only a few of the (predicted) differences between the eldest and youngest students in PISA are statistically significant. In particular, the eldest students in Albania tend to have higher socio-economic status than the youngest students (possibly due to drop-out of disadvantaged students around the age of 16), while the opposite is observed in Belarus. In Hungary, there are significantly fewer girls among the eldest students, compared to the youngest students. While such differences are controlled for in all remaining regressions included in this paper, their existence may reveal the possibility that in the above-mentioned countries, the results may be confounded by other systematic differences between the eldest and youngest students in PISA.

Table 5.4. Robustness check: Absence of selection effects associated with students' month of birth

Country/economy and years	Pseudo-grade-and-age effect[1] (% difference)						Number of observations
	High ESCS[2]		Gender: Girl		Immigrant background		
Albania (2018)	5.9	(3.7)	1.4	(2.7)	0.0	(0.4)	6 202
Baku (Azerbaijan)[3] (2018)	5.0	(3.4)	6.2	(3.3)	0.4	(1.8)	5 805
Belarus (2018)	-4.8	(2.1)	0.3	(3.1)	-0.9	(1.0)	5 631
B-S-J-Z (China) (2018)	1.0	(4.0)	0.1	(3.4)	-0.1	(0.3)	11 950
Costa Rica[3] (2015, 2018)	3.2	(3.3)	3.2	(2.8)	-2.3	(1.5)	12 397
Croatia (2015, 2018)	0.4	(2.1)	-0.6	(1.6)	-0.8	(1.2)	12 093
Czech Republic[3] (2015, 2018)	2.5	(2.5)	-4.2	(3.0)	0.8	(1.1)	10 004
Estonia (2015, 2018)	0.9	(2.1)	3.1	(2.8)	-1.3	(1.5)	10 610
Finland (2015, 2018)	-3.2	(2.0)	2.1	(1.6)	-1.5	(1.0)	11 285
Germany[3] (2015, 2018)	-5.0	(3.6)	1.4	(3.4)	2.2	(2.7)	6 621
Hong Kong (China) (2015, 2018)	3.9	(2.7)	1.3	(2.3)	0.6	(2.1)	10 920
Hungary (2015, 2018)	4.2	(4.3)	-1.2	(3.9)	-0.1	(1.3)	10 604
Korea (2018)	-0.8	(3.7)	2.0	(4.1)	-0.1	(0.6)	6 568
Luxembourg (2015, 2018)	0.9	(3.1)	-0.5	(3.3)	-3.6	(2.9)	10 093
Serbia (2018)	2.2	(3.4)	-0.2	(3.5)	-3.1	(1.7)	6 378
Slovak Republic[3] (2015, 2018)	-0.2	(2.6)	1.2	(2.7)	-0.6	(0.6)	8 700
Switzerland[3] (2015, 2018)	-2.4	(4.8)	-5.4	(5.0)	4.6	(4.9)	7 573
Chinese Taipei (2015, 2018)	-2.2	(2.0)	2.1	(2.3)	0.2	(0.2)	14 736

Notes: Each row corresponds to a separate regression. All regressions include only grade and school-starting-age variables, and, for countries/economies whose estimates are based on the pooled 2015 and 2018 data, a year dummy. Standard errors that account for clustering and for the sampling design are presented in parentheses.

1. Pseudo-grade-and-age effects for each variable are estimated using separate instrumental-variable regressions. They correspond to the coefficient on the grade variable (instrumented by the theoretical grade) and are reported as percentage-point differences. See Equation 4.1 for details.
2. High ESCS refers to students in the top half of the country's distribution of the index of economic, social and cultural status.
3. The analysis does not include students born in certain months. See Section 4 for details.

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

5.3. How transversal competences develop over a school year

In addition to measuring students' proficiency in three foundational domains of competence – reading, mathematics and science – which correspond to traditional subject areas of school curricula around the world, PISA also offers an innovative assessment developed expressly for each new survey cycle. The most recent innovative assessments were collaborative problem solving (2015) and global competence (2018). The PISA 2015 collaborative problem solving assessment measured students' capacity to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills and efforts to reach that solution (OECD, 2017[26]). The PISA 2018 global competence assessment measured students' capacity to examine local, global and intercultural issues, to engage in open, appropriate and effective interactions with people from different cultures, and to act for collective well-being and sustainable development (OECD, 2020[27]).

Both collaborative problem solving and global competence are not directly related to a particular school subject and can be seen as examples of transversal skills, which develop and can be used in a variety of settings. The scale on which the results of the collaborative problem solving and global competence tests are reported has similar properties as the scales for mathematics, reading and science; in particular, the standard deviation of the scale, which summarises the variation of results within each country and economy, is typically slightly below 100 score points. Thus, although it is not possible to compare the scores and score-differences across subjects directly because they are not expressed in a common unit, these score differences roughly correspond to effect-size measures. To this end, Table 5.5 reports the mean and standard deviation of performance in collaborative problem solving and global competence for all countries and economies included in the analysis.

Table 5.5. Mean score and variation in performance in collaborative problem solving and global competence in 18 countries and economies

Country/economy	Collaborative problem solving (PISA 2015)		Global competence (PISA 2018)	
	Mean	Standard deviation	Mean	Standard deviation
Albania	m	m	427	78
Baku (Azerbaijan)	m	m	m	m
Belarus	m	m	m	m
B-S-J-Z (China)	m	m	m	m
Costa Rica	441	78	456	86
Croatia	473	87	506	90
Czech Republic	499	91	m	m
Estonia	535	90	m	m
Finland	534	101	m	m
Germany	525	101	m	m
Hong Kong (China)	541	90	542	97
Hungary	472	95	m	m
Korea	538	84	509	96
Luxembourg	491	100	m	m
Serbia	m	m	463	99
Slovak Republic	463	93	486	97
Switzerland	m	m	m	m
Chinese Taipei	527	90	527	92

Note: m: missing (the country/economy did not participate in the corresponding assessment).
Source: OECD PISA 2015 and 2018 datasets, https://www.oecd.org/pisa/data/ (accessed on 17 May 2021).

Table 5.6 reports the effect of one year of schooling and age on performance in collaborative problem solving and in global competence. The variation, across countries, in the estimated grade gain in collaborative problem solving is strongly related to the variation in the grade gain for mathematics (r = 0.69) and science (r = 0.66) but relates more weakly to the grade gain in reading (r = 0.30). In contrast, the variation in the grade gain for global competence is most strongly related to the variation in the grade gain in reading (r = 0.92), but less so to the variation in the grade gain in mathematics (r = 0.61). This may suggest that collaborative problem-solving skills are often developed at school through learning experiences – such as group work – in scientific disciplines, whereas global competence is developed mostly through reading activities or activities that contribute, also, to the development of reading competencies.

The results also show, for most countries, somewhat smaller effect sizes for collaborative problem solving compared to reading and mathematics. The difference is particularly wide for the Czech Republic, Luxembourg and the Slovak Republic, suggesting that in these countries, there may be a limited emphasis around the age of 15 on developing transversal skills that can sustain collaboration. The main exception to this pattern is Chinese Taipei, where the grade-and-age effect on collaborative problem solving (17 score points) appears to be larger than in mathematics and reading (11 and 7 score points, respectively).

Table 5.6. Age-and-grade effects on collaborative problem solving and global competence

Country/economy	Collaborative problem solving (2015)					Global competence (2018)				
	Grade-and-age effect[1]		First stage: Effect of the expected grade on the actual grade[2]		Number of observations	Grade-and-age effect[1]		First stage: Effect of the expected grade on the actual grade[2]		Number of observations
Albania	m	m	m	m	m	8.3	(6.8)	0.82	(0.02)	6 202
Baku (Azerbaijan)	m	m	m	m	m	m	m	m	m	m
Belarus	m	m	m	m	m	m	m	m	m	m
B-S-J-Z (China)	m	m	m	m	m	m	m	m	m	m
Costa Rica[3]	20.0	(8.5)	0.55	(0.03)	5 972	26.7	(6.3)	0.69	(0.02)	6 425
Croatia	14.2	(5.8)	0.85	(0.01)	5 617	15.8	(6.4)	0.84	(0.01)	6 476
Czech Republic[3]	2.0	(8.3)	0.91	(0.02)	4 917	m	m	m	m	m
Estonia	29.1	(7.3)	0.79	(0.02)	5 447	m	m	m	m	m
Finland	16.9	(8.1)	0.98	(0.01)	5 782	m	m	m	m	m
Germany[3]	26.3	(10.5)	0.69	(0.02)	3 600	m	m	m	m	m
Hong Kong (China)	18.9	(7.2)	0.85	(0.02)	5 140	27.4	(8.5)	0.84	(0.03)	5 780
Hungary	-1.3	(14.7)	0.44	(0.02)	5 543	m	m	m	m	m
Korea	c	c	c	c	c	16.2	(7.2)	0.88	(0.01)	6 568
Luxembourg	20.0	(8.5)	0.55	(0.03)	5 085	m	m	m	m	m
Serbia	m	m	m	m	m	23.7	(8.6)	0.67	(0.02)	6 378
Slovak Republic[3]	10.2	(8.3)	0.89	(0.02)	4 414	38.6	(7.6)	0.87	(0.02)	4 286
Switzerland	m	m	m	m	m	m	m	m	m	m
Chinese Taipei	17.0	(5.8)	0.99	(0.00)	7 615	7.2	(5.6)	0.99	(0.01)	7 121

Notes: Each line represents separate regressions. All regressions include, in addition to grade and school-starting-age variables, controls for students' socio-economic status (three dummies, for quarters), gender and immigrant background.
1. Grade-and-age effects are estimated using instrumental-variable regressions. They correspond to the coefficient on the grade variable (instrumented by the theoretical grade). See Equation 4.1 for details.
2. The first-stage regression also includes age and other exogenous variables (control variables), whose coefficients are not reported.
3. The analysis does not include students born in certain months. See Section 4 for details.
m: missing (the country/economy did not participate in the corresponding assessment); c: there are no observations to estimate Equation 4.1 (all students are expected to be in the same grade).
Source: OECD PISA 2015 and 2018 datasets, https://www.oecd.org/pisa/data/ (accessed on 17 May 2021).

6. Discussion and interpretation

This final section discusses the interpretation of the estimates of grade gains presented in previous sections as well as their implications for policy and research.

6.1. Interpretation of grade-and-age effects

Throughout this paper, the quantity of interest has been described as the combined effect of being older by one year and having completed an additional grade level. Indeed, given that all students are tested in PISA within a short testing window, the methods presented earlier do not allow for distinguishing the two age effects (school-starting-age and age-at-testing effects) and for disentangling age-at-testing effects from the "pure" grade effect resulting from additional schooling. The distinction, however, may be relevant when discussing, for example, the potential impact of school closures in 2020 and 2021. Even during periods of school closure and in the absence of any form of school instruction, maturity effects (which result from physiological ageing and from the accumulation of out-of-school experiences) would continue to contribute to skill acquisition.

Previous literature does not provide much guidance regarding the relative importance of school instruction and out-of-school experiences on skill acquisition. To answer this question, several studies in the United States have examined seasonal patterns in test scores, and some of these have highlighted a "summer learning loss". However, the finding of an average fall in test scores during the summer break in elementary school, reported in an influential meta-analysis of early studies (Cooper et al., 1996[28]), may not reflect a real loss of skill, but rather the use of different test forms, which are only imperfectly linked, to measure reading and mathematics ability at the end of a school year and at the beginning of the next school year (von Hippel and Hamrock, 2019[29]). 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[29]). 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 is decreasing as students move from elementary to lower secondary grades (Atteberry and McEachin, 2020[30]).

This literature, moreover, is largely limited to the United States. But evidence about the elementary-school years in the United States may not apply to 15-year-old students in a much broader set of countries; and the contribution of school instruction and out-of-school experiences to the skills measured in a non-curricular test, such as PISA, may differ from their contribution to skills measured in tests more closely aligned to the school curriculum.

The interpretation of the estimates in this paper must therefore remain tentative. If school instruction is the most important factor influencing the acquisition of reading, mathematics and science skills measured by PISA (as would be implied by a "summer learning loss"), then the estimates reflect mostly "pure" grade effects, and periods with no instruction will likely see stagnation or even decline in PISA test scores. If, on the other hand, school instruction complements and adds to other experiences through which students learn and practice these skills, then the estimates reflect both grade and age-at-testing effects, and periods with no instruction will (at most) see a slowdown in test gains.

6.2. Variation in grade-and-age effects across countries and comparison of estimates with other studies

On average across PISA participants included in the present study, the grade gain in all three domains is between 18.8 and 19.9 score points (see Table 6.1), or about one-fifth of a standard deviation.

Table 6.1. Summary of grade and age effects in PISA

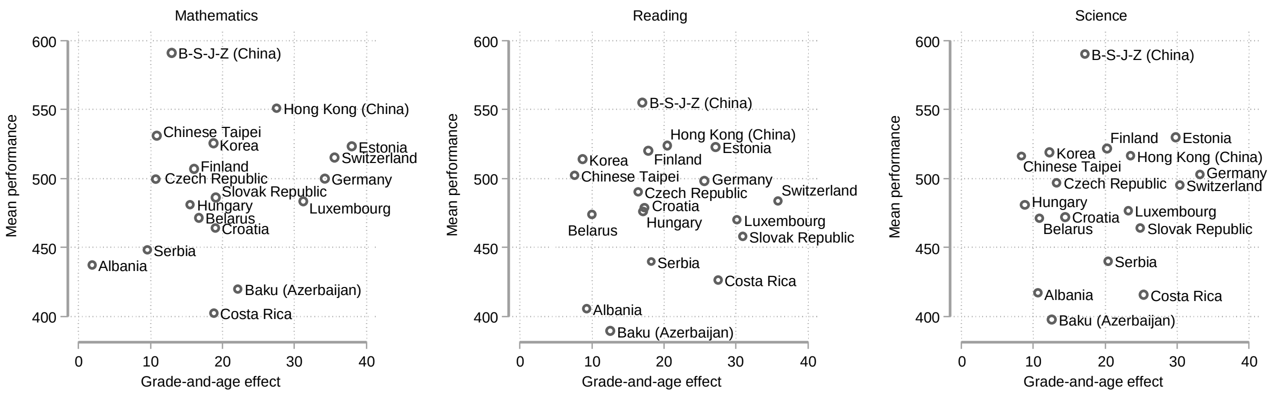
Meta-analytic summary of results

Domain	Countries and economies	Grade-and-age effects				
		Average	(Standard error)	Standard deviation	Minimum	Maximum
Mathematics	18	19.9	(1.5)	9.9	1.9	37.9
Reading	18	19.4	(1.4)	8.4	7.6	35.8
Science	18	18.8	(1.5)	7.9	8.4	33.1

Note: The average effect is computed as the arithmetic average across countries/economies. The standard error for the average effect is computed by exploiting the independence of samples across countries/economies.
Source: Table 5.1 in this paper.

Some countries and economies with large estimated grade gains, such as Estonia, also have high average scores in PISA; but overall, the association between grade-and-age effects and mean performance is weak (linear correlations range between 0.06 in reading and 0.22 in mathematics) (Figure 6.1). The weak correlation may be due to the large statistical uncertainty around grade-gain estimates, which results in attenuation bias. At the same time, the comparatively small grade gains for some high-performing countries/economies could indicate that in those countries and economies, results on international assessments of student learning reflect mostly an advantage built in the early grades, but that productivity decreases in the transition to upper secondary education. For example, the Czech Republic, Finland, Korea and Chinese Taipei were among the highest-performing countries and economies in TIMSS in 2011 among fourth-grade students – when the assessed cohort was, typically, in between the cohorts assessed in PISA 2015 and 2018 (Mullis et al., 2012[31]).

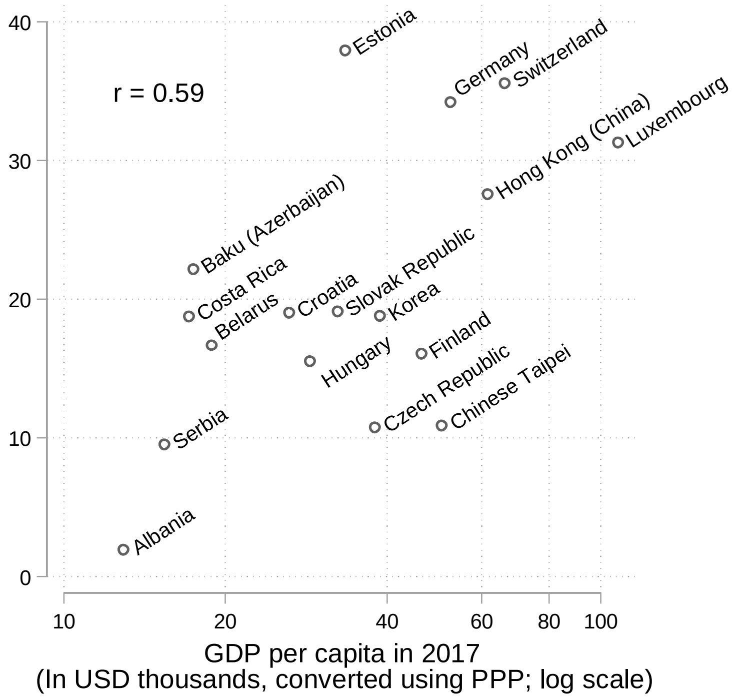
Figure 6.1. Grade effects and mean scores in PISA



Source: Table 3.3 and Table 5.1 in this paper.

A somewhat stronger association is observed between the estimated grade gains and gross domestic product (GDP) per capita. In particular, (log) GDP per capita shows a moderate correlation with the grade gain in mathematics (r = 0.59), reading (r = 0.43) and science (r = 0.39) (correlations based on 17 countries and economies, excluding B-S-J-Z [China]) (Figure 6.2).

Figure 6.2. Grade effects in PISA and GDP per capita in 17 countries and economies



Notes: B-S-J-Z (China) is not included in this figure due to the absence of GDP data. The GDP figure for Baku (Azerbaijan) refers to all of Azerbaijan.
Source: Table 5.1 in this paper and International Monetary Fund (2019[32]), *World Economic Outlook Database*, https://www.imf.org/en/Publications/WEO/weo-database/2019/April (accessed on 28 June 2021).

The "local" nature of the average grade gain identified using an instrumental-variable strategy may also explain some of the variation observed across countries. In particular, the grade gain reported in this paper should be interpreted as the average grade gain across all students for whom the instrument – the expected grade level, given school-entry regulations – is associated with an actual variation in the grade level ("compliers"). This population excludes all students who would have been enrolled in a higher or lower grade, irrespective of their birthday. In Costa Rica, Germany, Luxembourg and Switzerland, whose average grade gains are larger (but not always statistically significantly so) than those of other countries with similar mean performance, the population of compliers is more restricted, as indicated by lower first-stage coefficients in Table 5.1. Because non-compliers are mostly grade repeaters, whose learning gains in any given grade can be expected to be smaller than those of non-grade repeaters, it is plausible that the average learning gain for compliers is somewhat larger than the average learning grade for the full PISA cohort.

For a few countries and economies, it is possible to compare the grade-and-age effects estimated in the present study (Table 5.1), which relies on instrumental variables and a linear specification of age effects, with those obtained using different samples and estimation methods. In particular, Avvisati and Givord (2021[19]) exploit the variation in testing dates in the early cycles of PISA to estimate the grade gain in reading in Hong Kong (China) under less restrictive assumptions about age effects. While the two estimates refer to different years, the difference between the two point estimates (20.4 score points in the present study; 25.6 score points in the former) is not statistically significant.[16] Another benchmark is provided, for Germany, by longitudinal studies that followed students over a complete year of schooling and administered a second PISA test at one year of interval to the same students. A first German study (based on the PISA 2003 cohort) showed that over a one-year period (which corresponds both to a different age and a different grade), students gained about 25 score points in the PISA mathematics test, on average, and progressed by a similar amount (21 points) in a test of science (Prenzel et al., 2006[21]) – a result that is smaller, but not significantly so, than the estimates presented in this paper (34 and 33 points, respectively, in mathematics and science).[17]

Estimates of the grade gain that refer to different subjects are in general highly correlated, meaning that countries with a large estimated grade gain in one subject tend to have a similarly large grade gain in the two remaining subjects too. All pairwise linear correlations of country-level estimates across subjects are above 0.5; the weakest correlation is between mathematics and reading, 0.70, and the strongest correlation between reading and science, 0.87. This is not surprising, given the high correlation of PISA scores at the student level. While occasionally differences in point estimates across pairs of countries and economies are in the opposite direction, suggesting that students progress relatively more rapidly in one subject and more slowly in another subject, the statistical uncertainty associated with point estimates does not allow such conclusions to be drawn with confidence.[18]

[16] The significance of the difference can be estimated by exploiting the independence of samples upon which they are based, which means that the sampling covariance between the two estimates is exactly 0.

[17] The use of a longitudinal design to estimate trajectories of test-score growth presents an intuitive appeal, due to its reliance on within-student comparisons, but also considerable logistic challenges, and is not exempt from significant threats to validity. These are related to the possible selection bias that results from attrition in follow-up waves; to the possible confounding role of test design and test-administration conditions, and of student effort and motivation, on measurement invariance and on test results. The difficulty of interpreting within-student comparisons of test scores at one year of interval as "grade-and-age effects" are illustrated by the second longitudinal PISA study in Germany, based on the PISA 2012 cohort. This study was affected by selective attrition (Heine et al., 2017[41]) and significant test-motivation effects, and, after controlling for these (based on strong assumptions), found much smaller average performance "gains" than in the previous study. In fact, the learning gains were not significantly different from 0 in reading and in science (Nagy et al., 2017[36]).

[18] It is not possible to compare the scores in reading, mathematics and science directly because they are not expressed in a common unit; but it is possible to compare two countries with each other across all subjects, taking into account the statistical uncertainty associated with point estimates. When performing such pairwise comparisons of vectors of coefficients, the simultaneous testing of multiple hypotheses (one per subject) is accounted for by simulation procedures that consider the positive dependence of the estimated coefficients across subjects, within countries. In particular, 10 000 replicate values of subject-specific grade effects are generated, according to multi-variate normal distributions, where the covariance structure reflects the observed covariance of subject-specific estimates across countries. This simulated dataset is used to identify country/economy pairs for which one can reject, at the 5% significance level, null hypotheses of the form $H_0: \beta_{c,s} \geq \beta_{c',s}\ or\ \beta_{c,s'} \leq \beta_{c',s'}$, where c and c' are two countries and s and s' two subjects (the alternative hypothesis is $\beta_{c,s} < \beta_{c',s}\ and\ \beta_{c,s'} > \beta_{c',s'}$). Based on Montecarlo simulations, the null at the 5% level can never be rejected.

6.3. Using grade gain estimates as a benchmark

The estimates reported in the present paper, based on quasi-experimental methods, indicate that for 15-year-old students, the typical grade gain is about one-fifth of a standard deviation, or around 19 score points. When projecting the long-term economic impact of school closures due to COVID-19, Hanushek and Woessmann (2020[33]) assume that PISA scores increase on average by one-third of a standard deviation over a school year,[19] i.e. about 33 score points. Psacharopoulos et al. (2020[34]) equate a year of schooling with a difference of about 40 score points in upper-middle-income countries (with a range from 20 to 50 points depending on the country's income level), relying on "naïve" estimates of grade effects that exploit the variation in grades observed in PISA and treat it as exogenous (Azevedo et al., 2020, p. 41[35]). The evidence about grade-and-age effects in the present paper therefore implies that (potential) economic losses due to school closures may be only about half as big as initially projected by these authors. At the same time, these authors assume, in their projections, that PISA scores increase by a similar extent at every grade level; the fact that the grade gain at age 15 is smaller than was assumed does not automatically imply that the grade gain is smaller at every other grade level, and in particular in lower grades.

The average grade effect for 15-year-olds estimated in the present paper (around 20 score points) can also 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 Canada (512 points) was larger than the typical test-score gap between students who are one grade level apart, around the age of 15; as was the average gender gap in reading (30 score points, on average across OECD countries). And it would take students in the bottom 25% of socio-economic status, who score on average 89 points lower than students in the top 25% in reading, several years of schooling to reach the current level of their more advantaged peers. 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.

[19] "A rough rule of thumb, found from comparisons of learning on tests designed to track performance over time, is that students on average learn about one-third of a standard deviation per school year" (Hanushek and Woessmann, 2020, p. 10[33]).

References

Almond, D. and B. Mazumder (2011), “Health Capital and the Prenatal Environment: The Effect of Ramadan Observance During Pregnancy”, *American Economic Journal: Applied Economics*, Vol. 3/4, pp. 56-85, http://dx.doi.org/10.1257/app.3.4.56. [47]

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. [9]

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. [30]

Avvisati, F. and P. Givord (2021), “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 Papers*, No. 249, OECD Publishing, Paris, https://dx.doi.org/10.1787/d99e8c0a-en. [19]

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. [48]

Avvisati, F. and N. Viltard (forthcoming), “How much do the skills of Italian 15-year-olds improve over one year of schooling?”, in *Volume 5 - The School and its Protagonists: The Students*, Franco Angeli. [22]

Azevedo, J. et al. (2020), *Simulating the Potential Impacts of COVID-19 School Closures on Schooling and Learning Outcomes: A Set of Global Estimates*, The World Bank, http://dx.doi.org/10.1596/1813-9450-9284. [35]

Barua, R. and K. Lang (2016), “School Entry, Educational Attainment, and Quarter of Birth: A Cautionary Tale of a Local Average Treatment Effect”, *Journal of Human Capital*, Vol. 10/3, pp. 347-376, http://dx.doi.org/10.1086/687599. [49]

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. [13]

Buckles, K. and D. Hungerman (2013), “Season of Birth and Later Outcomes: Old Questions, New Answers”, *Review of Economics and Statistics*, Vol. 95/3, pp. 711-724, http://dx.doi.org/10.1162/rest_a_00314. [45]

Cáceres-Delpiano, J. and E. Giolito (2018), “Minimum Age Requirements and the Impact of School Choice”, *Discussion Paper Series*, No. 11420, IZA, http://ftp.iza.org/dp11420.pdf (accessed on 24 September 2020). [40]

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. [28]

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. [12]

Dhuey, E. (2016), “Age at school entry: How old is old enough?”, *IZA World of Labor*, http://dx.doi.org/10.15185/izawol.247. [14]

Engzell, P., A. Frey and M. Verhagen (2021), “Learning loss due to school closures during the COVID-19 pandemic”, *Proceedings of the National Academy of Sciences*, Vol. 118/17, p. e2022376118, http://dx.doi.org/10.1073/pnas.2022376118. [3]

Fernández Aráuz, A. (2016), *Educación Preescolar en Costa Rica: Historia y Situación Actual*, Ministerio de Educación Pública, San José, Costa Rica, https://www.mep.go.cr/indicadores_edu/BOLETINES/Preescolar.pdf (accessed on 7 October 2020). [44]

Gans, J. and A. Leigh (2009), “Born on the first of July: An (un)natural experiment in birth timing”, *Journal of Public Economics*, Vol. 93/1-2, pp. 246-263, http://dx.doi.org/10.1016/j.jpubeco.2008.07.004. [17]

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. [24]

Hanushek, E. and L. Woessmann (2020), “The economic impacts of learning losses”, *OECD Education Working Papers*, No. 225, OECD Publishing, Paris, https://dx.doi.org/10.1787/21908d74-en. [33]

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. [41]

Huang, C., S. Zhang and Q. Zhao (2020), “The early bird catches the worm? School entry cutoff and the timing of births”, *Journal of Development Economics*, Vol. 143, p. 102386, http://dx.doi.org/10.1016/j.jdeveco.2019.102386. [18]

International Monetary Fund (2019), *World Economic Outlook Database*, https://www.imf.org/en/Publications/WEO/weo-database/2019/April (accessed on 28 June 2021). [32]

INVALSI (2021), *Rilevazioni nazionali degli apprendimenti 2020-21: I risultati in breve delle prove INVALSI 2021*, https://invalsi-areaprove.cineca.it/docs/2021/Rilevazioni_Nazionali/Rapporto/14_07_2021/Sintesi_Primi_Risultati_Prove_INVALSI_2021.pdf (accessed on 4 August 2021). [6]

Kim, T. (2021), “Age culture, school-entry cutoff, and the choices of birth month and school-entry timing in South Korea”, *Journal of Demographic Economics*, Vol. 87/1, pp. 33-65, http://dx.doi.org/10.1017/dem.2020.16. [15]

Kuhfeld, M. et al. (2020), *Learning during COVID-19: Initial findings on students' reading and math achievement and growth*, NWEA, https://www.nwea.org/content/uploads/2020/11/Collaborative-brief-Learning-during-COVID-19.NOV2020.pdf (accessed on 5 May 2021). [4]

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. [8]

Liu, D. and X. Li (2016), "Children born in July and August: a study on the age regulation in primary school and student's education access and development", *The Journal of Chinese Sociology*, Vol. 3/1, http://dx.doi.org/10.1186/s40711-016-0042-1. [42]

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. [7]

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. [11]

Majid, M. (2015), "The persistent effects of in utero nutrition shocks over the life cycle: Evidence from Ramadan fasting", *Journal of Development Economics*, Vol. 117, pp. 48-57, http://dx.doi.org/10.1016/j.jdeveco.2015.06.006. [46]

Maldonado, J. and K. De Witte (2020), *The Effect of School Closures on Standardised Student Test Outcomes*, https://lirias.kuleuven.be/retrieve/588087. [5]

Mullis, I. et al. (2012), *Timss 2011 International Results in Mathematics*, TIMSS & PIRLS International Study Center and International Association for the Evaluation of Educational Achievement (IEA), https://timssandpirls.bc.edu/timss2011/international-results-mathematics.html. [31]

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. [36]

Neugart, M. and H. Ohlsson (2012), "Economic incentives and the timing of births: evidence from the German parental benefit reform of 2007", *Journal of Population Economics*, Vol. 26/1, pp. 87-108, http://dx.doi.org/10.1007/s00148-012-0420-1. [16]

OECD (2020), *PISA 2018 Results (Volume VI): Are Students Ready to Thrive in an Interconnected World?*, PISA, OECD Publishing, Paris, https://dx.doi.org/10.1787/d5f68679-en. [27]

OECD (2017), "PISA 2015 collaborative problem-solving framework", in *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematics, Financial Literacy and Collaborative Problem Solving*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264281820-8-en. [26]

OECD (2013), *PISA 2012 Results: What Makes Schools Successful (Volume IV): Resources, Policies and Practices*, PISA, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264201156-en. [37]

OECD (2010), *PISA 2009 Results: What Makes a School Successful?: Resources, Policies and Practices (Volume IV)*, PISA, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264091559-en. [38]

OECD (2007), *PISA 2006: Science Competencies for Tomorrow's World: Volume 1: Analysis*, PISA, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264040014-en. [39]

Prenzel, M. et al. (eds.) (2006), *PISA 2003: Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*, Waxmann Verlag GmbH. [21]

Psacharopoulos, G. et al. (2020), "Lost Wages : The COVID-19 Cost of School Closures", *Policy Research Working Paper*, No. 9246, World Bank, https://openknowledge.worldbank.org/handle/10986/34387. [34]

Renaissance Learning, Education Policy Institute (2021), *Understanding progress in the 2020/21 academic year: Interim Report*, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/962330/Learning_Loss_Report_1A_-_FINAL.pdf. [2]

Rose, S. et al. (2021), *Impact of school closures and subsequent support strategies on attainment and socio-emotional wellbeing in Key Stage 1: Interim Paper 1*, NFER, https://educationendowmentfoundation.org.uk/public/files/Publications/Covid-19_Resources/Impact_of_school_closures_KS1_interim_findings_paper_-_Jan_2021.pdf (accessed on 5 May 2021). [1]

Rubin, D. (1974), "Estimating causal effects of treatments in randomized and nonrandomized studies", *Journal of Educational Psychology*, Vol. 66/5, pp. 688-701, http://dx.doi.org/10.1037/h0037350. [23]

Sacerdote, B. (2011), "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?", in *Handbook of the Economics of Education*, Elsevier, http://dx.doi.org/10.1016/b978-0-444-53429-3.00004-1. [25]

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. [20]

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. [10]

Tourangeau, R. and N. Bradburn (2010), "The psychology of survey response", in Marsden, P. and J. Wright (eds.), *Handbook of Survey Research*, Emerald. [50]

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. [29]

Zhang, C. and Y. Xie (2018), “Does the school cutoff date cause the disadvantage for children born in July and August?”, *The Journal of Chinese Sociology*, Vol. 5/1, http://dx.doi.org/10.1186/s40711-018-0075-8. [43]