RESEARCH

Open Access

Equal opportunities for all? Analyzing within-country variation in school effectiveness

Isa Steinmann[*] and Rolf Vegar Olsen

*Correspondence:
isa.steinmann@cemo.uio.no
Centre for Educational
Measurement (CEMO),
University of Oslo, Blindern,
Postboks 1161, 0318 Oslo,
Norway

Abstract

Background: The present study addresses the equality of school effectiveness across schools. One central aim of compulsory education is for students to learn equally well at all schools in a country even if these schools differ in terms of student composition. However, measuring equality of learning gains independently from selection effects usually requires longitudinal data. This study suggests a different approach and proposes a new measure for the equality of school effectiveness across schools.

Methods: We applied a two-level regression discontinuity approach to estimate the between-school variation in added-year effects on mathematics and science achievement that result from an additional year of schooling, after controlling for the effects of age and student selection—i.e., between-school differences in achievement levels. We utilized data from a total of 13 samples. These stemmed from Nordic and other European countries, which assessed students from two adjacent grades at the same schools. The samples stemmed from TIMSS 1995 and 2015 and both primary and secondary school levels.

Results: The main findings indicated that although schools differed in initial achievement levels in all samples, schools in some countries, such as Norway and Cyprus, attained a high degree of equality of school effectiveness—i.e., of the effect of an additional year of schooling. Despite the fact that schools with a more privileged student composition had higher achievement levels than less privileged schools, their school effectiveness did not usually differ significantly. Both age and an additional year of schooling resulted in positive effects on mathematics and science achievement, however, effect sizes differed considerably between the 13 samples.

Conclusions: We discuss the implications of the proposed school effectiveness measure, which is based on a regression discontinuity approach. We conclude that countries, such as Nordic ones, should consider extending their participation in international large-scale assessments with additional grades from the same schools in future cycles. This design would enable a multitude of robust school effectiveness studies in the future.

Keywords: Regression discontinuity, Mathematics achievement, Science achievement, Educational inequality, School effectiveness, Socioeconomic composition, Age/grade effects

There are numerous reasons why students have differing learning opportunities. The students' socioeconomic background is among the most important ones—with children of highly educated, well-earning parents also being privileged in terms of educational achievement and attainment (Jerrim & Macmillan, 2015; Sirin, 2005; Strietholt et al., 2019). Family background disparities are, by implication, very difficult to overcome for both education stakeholders and education systems. However, one aim of compulsory primary and secondary education is to foster learning for all students, independent of their family backgrounds (Jackson, 2013; Marks, 2005; Schütz et al., 2008). Especially the social-democratic Nordic countries aim for their schools to provide the same learning opportunities to all students and to equally foster learning outcomes across schools (Blossing et al., 2014; Frønes et al., 2020; Yang Hansen et al., 2014).
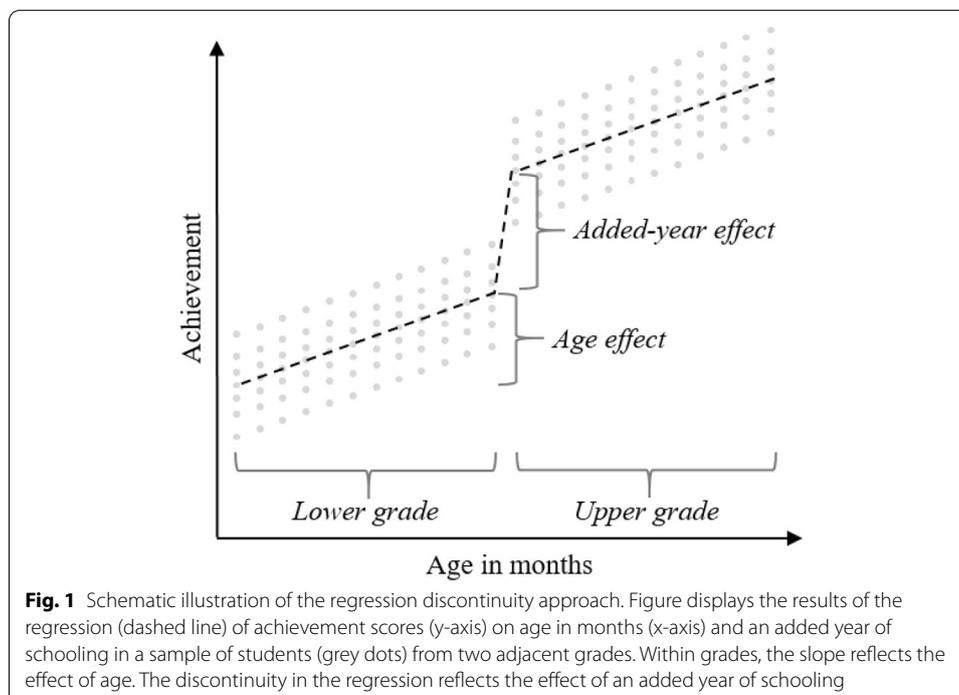
If an education system did, indeed, succeed in fulfilling the goal of providing the same learning opportunities at all schools, then school effectiveness measures should not vary between schools. Clearly, the ambition to have equally effective schools does not simply mean that all schools have the same mean achievement outcomes. Different student enrollment policies typically result in different student compositions across schools. Regardless of whether these policies relate to geographical catchment areas, merit-based selection, or more direct monetary mechanisms, the likely result is that schools will end up with different socioeconomic student compositions. This is important because the socioeconomic composition of the student body is associated with the students' initial mean achievement levels (Mullis et al., 2017, 2020; OECD, 2019b). Instead, equality of school effectiveness implies that students should make the same learning progress, independent of their school's socioeconomic composition. Hence, ideally, the learning gains per school year—a measure for school effectiveness—should be the same across all schools in a given country even if they do vary in terms of their student bodies (i.e., student selection). We consider equality of school effectiveness to be an important facet of educational effectiveness because it depicts whether some students are left behind at their schools (Bosker & Scheerens, 1994; Chapman et al., 2012; Townsend, 2007). This is also relevant information for policymaking because the policies that affect the allocation of highly qualified teachers and adequate resources for all schools, as well as other school-level policies, can more easily be changed through political reforms than, for instance, the socioeconomic backgrounds of students.

One way to disentangle school effectiveness from student selection effects at the school level is to use longitudinal studies with large sample sizes. In such studies, the between-school differences in the learning progress found between measurement points can be isolated from the achievement differences that had already existed at the first point of measurement. Such large-scale longitudinal studies are available in a number of countries (see overview in Blossfeld et al., 2019). Some of these countries also extend their participation in cross-sectional international assessments with a longitudinal follow-up measurement (e.g., OECD, 2012; Prenzel et al., 2006). Others, such as the Nordic countries, are able to longitudinally connect central data registers of their entire populations with results from standardized achievement tests and teacher-set grades (see e.g., Ræder et al., 2020; Figlio et al., 2016). Without a doubt, such longitudinal designs provide unique research opportunities for tracing school effectiveness differences between schools. However, these studies have some key disadvantages. They are expensive and

time-consuming to carry out, they typically suffer from high attrition rates, and they do not allow for international comparisons. Furthermore, they do not allow for effects of schooling to be disentangled from effects of aging. Achievement gains over one school year reflect both the effects of an additional year of school and those of an additional year of age, both of which should typically have positive effects on achievement outcomes. However, aging effects in school children should not be regarded as a facet of school effectiveness. In contrast, a method that is able to disentangle school effectiveness from both pure age and selection effects is the regression discontinuity design (cf. Luyten et al., 2009).

Utilizing regression discontinuity designs to estimate the equality of school effectiveness

The basic principle of the regression discontinuity approach is that small, random differences in the so-called running or forcing variable make a pronounced difference in terms of treatment because the so-called scoring rule applies (Angrist & Pischke, 2009, 2015; Imbens & Lemieux, 2008). In the present study, this means that small differences in the birth date (running variable age) decide whether a student is placed in a lower grade (control group) or upper grade (treatment group) due to school entry regulations that have fixed birth-date-related thresholds (scoring rule). In Norway, for instance, December-born children enter school a whole year earlier than children who are born only one month later, in January, because of school entry policies. This is illustrated in Fig. 1, where the age in months determines whether students attend the lower or upper of two adjacent grades. In terms of the outcome—the achievement in a standardized test—both older age (running variable) and an added year of schooling (treatment) should have positive effects. Due to brain maturation processes, older children should perform better

Fig. 1 Schematic illustration of the regression discontinuity approach. Figure displays the results of the regression (dashed line) of achievement scores (y-axis) on age in months (x-axis) and an added year of schooling in a sample of students (grey dots) from two adjacent grades. Within grades, the slope reflects the effect of age. The discontinuity in the regression reflects the effect of an added year of schooling

on tests than their younger classmates (age effect). Children who go to school one year longer should outperform schoolmates in a lower grade due to the learning effects of formal education (added-year effect).

In the regression discontinuity approach, the outcome (achievement score) is regressed on both the treatment (added year of schooling) and the running variable (age in months). The discontinuity of this regression at the scoring rule threshold reflects the treatment effect, i.e., the added-year effect. In Fig. 1, this regression is depicted as a dashed line. This so-called sharp regression discontinuity approach can be applied if enough students comply with the scoring rule, i.e., attend the formally correct grade in accordance with their birth date and school entry regulations. This is the case in education systems that have strict school entry regulations as well as rare grade retentions and accelerations. Other prerequisites that must be met to apply regression discontinuity designs (Schochet et al., 2010; see Appendix A) are usually fulfilled when regressing achievement on age and grade (Luyten, 2006; Luyten et al., 2017).

Previous regression discontinuity findings on age and added-year effects

Numerous previous studies have applied the regression discontinuity approach in order to disentangle the effects of age from the effects of schooling on cognitive ability (Cahan & Cohen, 1989; Cahan et al., 2008; Jabr & Cahan, 2014, 2015; Kyriakides & Luyten, 2009; Luyten et al., 2017; Singh, 2020) and school achievement outcomes (Ali & Heck, 2012; Cahan & Davis, 1987; Cahan et al., 2008; Cliffordson, 2010; Crone & Whitehurst, 1999; Gormley & Gayer, 2005; Heck & Moriyama, 2010; Kyriakides & Luyten, 2009; Luyten, 2006; Luyten & Veldkamp, 2011; Luyten et al., 2008, 2009, 2017, 2020; Olsen & Björnsson, 2018; Perry, 2017; Van Damme et al., 2010; Webbink & Gerritsen, 2013). Other studies have not applied the above-discussed regression discontinuity approach but have employed similar strategies to disentangle age and schooling effects (Baltes & Reinert, 1969; Carlsson et al., 2012; Cascio & Lewis, 2006; Cliffordson & Gustafsson, 2008; Stelzl et al., 1995; Van de Vijver & Brouwers, 2009).

The findings of these studies can be summarized as follows. Both age and schooling usually have positive effects on scores in more general cognitive as well as academic achievement tests. However, the literature provides rich and nuanced information beyond this simple fact. First, effect sizes vary considerably between countries (Luyten, 2006; Luyten & Veldkamp, 2011; Singh, 2020; Webbink & Gerritsen, 2013). Moreover, based on studies that are repeated at regular intervals, the age effects fluctuate considerably over time within the same country (Olsen & Björnsson, 2018).
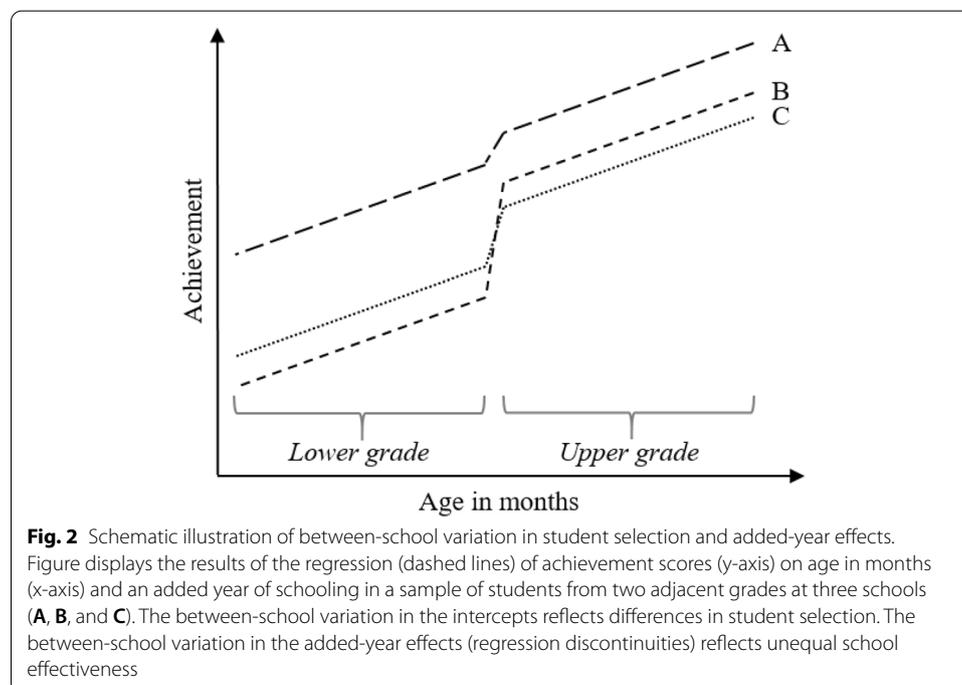
Second, effect sizes of age and added-year effects vary between younger and older students. Over their school career, both age and added-year effect sizes typically decrease (Cliffordson, 2010; Luyten et al., 2017; Olsen & Björnsson, 2018; Webbink & Gerritsen, 2013). However, this finding might also result from an increasing between-student variability in test scores, which would mean that absolute age and added-year effects are evaluated against a larger variance in older students. In one study with vertically linked scores, the schooling effect did not decrease over the school career (Kyriakides & Luyten, 2009).

Third, the literature indicates interesting differences between test domains. In comparison to tests on academic achievement domains that are closely tied to school curricula,

more general cognitive ability tests indicate smaller added-year effects (Cahan et al., 2008; Cliffordson & Gustafsson, 2008; Luyten et al., 2017). The obvious explanation for this finding is that schooling seems to have greater effects in areas that are explicitly taught at school (e.g., mathematics, reading) than on more general cognitive abilities (e.g., generic problem solving). Furthermore, two studies found that age effects on cognitive ability test scores decreased more over the school career than age effects on academic achievement test scores (Kyriakides & Luyten, 2009; Luyten et al., 2017). In terms of the added-year effect, no such differences between the two test domains were revealed (Luyten et al., 2017).

Extending the regression discontinuity approach to measure the equality of school effectiveness

The generic example of regression discontinuity in Fig. 1 provides a simple illustration of how this approach effectively disentangles the effects of age and schooling. By translating this approach to a multi-level framework, however, one can further separate the between-school differences in school effectiveness from student selection effects. This approach is illustrated in Fig. 2. Just as before, an outcome (achievement score) is regressed on the treatment (added year of schooling) and on the running variable (age in months). A two-level regression model includes a random intercept, a random slope for the added-year effect, and a fixed slope for the age effect. Hence, the model captures the differences in intercepts and added-year effects between schools, while the age effect is the same across schools. This is illustrated for three schools in Fig. 2. Example school A starts off with a higher achievement level in the lower grade than schools B and C (cf. intercepts in Fig. 2) but has a smaller added-year effect (cf. regression discontinuities

Fig. 2 Schematic illustration of between-school variation in student selection and added-year effects. Figure displays the results of the regression (dashed lines) of achievement scores (y-axis) on age in months (x-axis) and an added year of schooling in a sample of students from two adjacent grades at three schools (**A**, **B**, and **C**). The between-school variation in the intercepts reflects differences in student selection. The between-school variation in the added-year effects (regression discontinuities) reflects unequal school effectiveness

in Fig. 2). School B has the lowest achievement levels in the lower grade but the largest added-year effect. Age effects (cf. slopes in Fig. 2) are the same at all schools.

The two-level regression discontinuity approach allows for the estimation of the degree to which schools vary in their school effectiveness (i.e., the added-year effect), while controlling for the between-school differences in student selection (i.e., the intercept) that already preexisted in the lower grade. Although it is probably very difficult for countries to reduce the between-school variation in selection mechanisms (i.e., intercepts), which occur due to, for instance, residential segregation, countries can nevertheless aim to minimize the between-school variation in school effectiveness. Similarly, it is probably unavoidable for schools with a privileged student composition to have higher achievement intercepts (Mullis et al., 2017, 2020; OECD, 2019b). However, countries can aim to attain the same added-year slopes at all schools, regardless of their socioeconomic composition.

Previous two-level regression discontinuity findings

The variation in the added-year effects between schools, which serves as the central measure of interest in the present study, has rarely been investigated in the past. Only six of the above-mentioned regression discontinuity studies applied a multi-level extension to investigate the between-school variation in intercepts and added-year slopes.

One study that investigated age and added-year effects on mathematics and reading outcomes between grades 7 and 8 in 53 schools in the US found significant between-school variance in both intercepts and added-year slopes (Ali & Heck, 2012). At schools with a disadvantaged socioeconomic student composition, both intercepts and added-year slopes were lower than in schools with a more privileged student body. Another mathematics and reading domain study from the US also found significant between-school variations in both age and added-year effects in grade 4 and 5 students from 198 schools (Heck & Moriyama, 2010). In contrast to the first study, however, they found no significant correlation between the schools' added-year effects and socioeconomic compositions. However, both studies were based on complex approaches, where age and added-year effects were embedded into multi-level structural equation models with multiple predictors and interaction terms at student and school levels. Hence, the interpretation of the results is not straightforward.

A multi-level regression discontinuity study that used data from 15-year-old students from 270 schools in England found a significant between-school variation in intercepts but not in added-year effects (Luyten et al., 2008). In complex models with multiple predictors, schools with disadvantaged student compositions showed lower intercepts but higher added-year effects than schools with more privileged student bodies. Another study that used primary school level data from England found that 18 schools varied significantly in both intercepts and added-year effects (Luyten et al., 2009), but it did not investigate associations with school composition.

In a multi-level regression discontinuity study that used data from six secondary school grades at six Cypriot schools, the schools significantly differed in added-year effects on mathematics but not on Greek language or cognitive ability tests (Kyriakides & Luyten, 2009). The study, possibly due to the small sample of schools, found no significant association between added-year effects and any student background characteristics.

Another study used primary school level data from eight countries that participated with two adjacent grades in the Third International Mathematics and Science Study (TIMSS) in 1995 (Luyten, 2006). In all countries, both intercepts and added-year effects on mathematics and science test scores varied between schools, but this variation was not significant in all countries. The between-school variations of intercepts exceeded those of the added-year effects. Interestingly, the countries showed very different degrees of between-school variation of added-year effects. In none of the countries did the schools with a more privileged student body significantly differ from others in terms of added-year effects. However, these models contained numerous further variables at both the student and school levels, which again complicates the interpretability of their estimates.

In summary, existing studies showed that schools usually differ in intercepts internationally, i.e., achievement levels in lower grades. Furthermore, schools with a more privileged student composition were found to have higher intercepts than others. These findings are unsurprising because achievement intercepts reflect the fact that schools recruit from residential areas that are differently composed as well as because schools might have different school effectiveness levels in prior grades. In contrast, not all studies found pronounced between-school variation in added-year effects. While this might relate to the sometimes small school sample sizes, among other issues, it could also reflect the fact that some countries, indeed, manage to attain a very similar school effectiveness across schools. In most studies, school composition was not significantly associated with the schools' added-year effects. However, these studies included a multitude of additional predictor variables and different interaction terms at individual and school levels in their models, which may obscure and complicate the interpretability of their findings.

Advantages of the regression discontinuity approach for investigating the equality of school effectiveness

The previous literature has demonstrated that the multi-level regression discontinuity approach can be used to investigate the equality of school effectiveness by modeling the between-school variation of added-year effects. In comparison with alternative longitudinal designs, this approach has several advantages (Angrist & Pischke, 2009, 2015). First, longitudinal analyses are more demanding to carry out, both in terms of cost and the time that elapses before data can be analyzed. For the regression discontinuity approach, in contrast, student data of at least two adjacent grades can be collected at the same point in time. Second, the regression discontinuity approach does not suffer drawbacks from students changing schools and other sample attrition causes as longitudinal studies usually do. Third, it is much easier to repeat the same regression discontinuity study after a few years in order to obtain trend information than to repeat a longitudinal study. Fourth, it is much easier to conduct the same regression discontinuity study in multiple countries in order to enable a comparison of the equality of school effectiveness, as was done in TIMSS 1995. Unfortunately, the subsequently repeated TIMSS cycles (Trends in International Mathematics and Science Study) only targeted one grade per school in their international design. In addition, it is important to emphasize that studies that did directly compare the results of longitudinal and regression discontinuity

approaches found that these reached very similar conclusions at the country level (Luyten et al., 2009; Perry, 2017; Singh, 2020).

Nevertheless, the regression discontinuity approach is also associated with potential limitations. The most important one for the current study is that its design requires large complier proportions—i.e., students who are in the formally correct grade given their birth date and school entry regulations. In countries with strict school entry policies and low rates of grade repetition and acceleration, such as the Nordic countries, the regression discontinuity approach is a promising alternative to longitudinal designs for obtaining robust estimations of the inequality of school effectiveness.

The present study

This study aims to disentangle the between-school variation in school effectiveness from the between-school variation in student selection. We applied the above-described two-level regression discontinuity approach to estimate the between-school variation in added-year effects (school effectiveness) and intercepts (student selection). Furthermore, this study investigates whether the schools' added-year effects and intercepts differ depending on their socioeconomic student composition. This study uses Norway as a showcase with other carefully selected Nordic and European countries serving as a basis for comparison. Altogether, 13 samples are investigated—from 7 countries, 2 grade levels, and TIMSS 1995 and 2015—that were identified as being complicit with the standards for conducting regression discontinuity analyses. On this basis, the current paper demonstrates how the equality of school effectiveness measure can be compared between countries and over a 20-year interval. Furthermore, the study includes both primary and secondary school levels. However, their scores are not directly comparable because they were scaled independently. To complement previous research, the central strengths of this study lie in the fact that we ran simpler and more targeted models to identify the between-school variation in school effectiveness. In addition, the reanalyzed data from previous studies are complemented with an analysis of new data.

Methods
Data

The main data source of interest came from Norway's extended participation in TIMSS 2015, where the international sampling design was complemented by the assessment of 2 adjacent grades at the same schools and at both primary (grades 4 and 5) and secondary (grades 8 and 9) school levels. To compare these findings with other relatively similar countries and with an earlier point in time, we added 11 primary and secondary school samples from TIMSS 1995 in which 2 grades from the same schools were assessed. This study partly uses the same datasets as Luyten (2006). We applied a strict criterion for selecting countries from the TIMSS 1995 study in order to ensure that only samples with a high share of compliers were included: the proportion of students in the formally correct grades—given their birthdates and the countries' school entry regulations—was at least 93% in the samples included for analysis. In some of the countries excluded from the analysis, such as France, Germany, and Ireland, the shares of compliers were well below 80%. In addition, we tested the assumptions for regression discontinuity designs for all samples (see below).

Table 1 illustrates the sample preparation for the 13 selected samples. First, we identified those students with complete birth and test date information who attended schools that assessed two adjacent grades. Second, we identified the share of compliers (see columns 1–3 in Table 1). Third, we only included students from school samples with at least 10 lower and 10 upper grade students in order to enable analyses at the school level (see columns 4–11 in Table 1). Apart from the Norway 1995 secondary school sample, the number of students exceeded 2,000 and the number of schools 50 in the analyses samples. With the exception of the Sweden 1995 secondary school sample, the lower and upper grades contained approximately balanced numbers of students. It should be noted here that there are differences in grade labels and mean age (see columns 7, 8, 10, and 11 in Table 1) and that these reflect different policies for school starting ages in the selected countries.

Measures

Outcome variables

The outcome variables were student achievement scores in mathematics and science. In both TIMSS 1995 and 2015, students responded to standardized paper–pencil-based tests that contained multiple choice and open-ended items (Martin et al., 2016, Martin & Kelly, 1996). Two age-adequate test versions for primary and secondary school students were used, respectively. Per age group, students from the lower and upper grades responded to the same tests, and the achievement scores were scaled to allow for the comparison of lower and upper grade students' scores. As achievement scores, TIMSS provided five plausible values per student, which were computed using conditioning techniques with test and background data. Per age group and achievement domain, the scores had an international mean of 500 and standard deviation of 100 (Martin & Kelly, 1997; Martin et al., 2016). Plausible values contained no missing values.

Predictor variables

We used two main predictor variables—added year of schooling (treatment) and age (running variable). Per country sample, we recoded the grade variable to an added-year dummy variable for treatment (1 = upper grade) versus control (0 = lower grade) group membership. Age in months was determined based on the students' birth month and year relative to the test month and year. To facilitate the interpretability of intercepts in later regression models, we recoded the age in months relative to the threshold age between grades (0 = oldest students in lower grade, 1 = youngest student in upper grade) so that the age variable ranges between − 11 (youngest students in lower grade) and 12 (oldest students in upper grade). Since we only included students attending formally correct grades, none of the students in the analyses samples were formally too young or old for their grade. As mentioned above, students with missing test or birth date information were excluded because their formally correct grades could not be determined. In the original samples, the share of missing date values were in the 0–2% range, except for the Scotland 1995 primary school sample (11%). The grade variable contained no missing values.

Furthermore, we used the school-level aggregated number of books at students' homes as a proxy for socioeconomic school composition. In both TIMSS 1995 and

Table 1 Overview of countries and student samples

#	School level Country	Complier samples			Analysis samples							
							Lower grade			Upper grade		
		% Complier	# Students	# Schools	# Students	# Schools	# Students	Grade Label	Mean Age	# Students	Grade Label	Mean Age
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Primary school											
1	Norway 2015	99	8303	137	8069	124	3987	4	9.72	4082	5	10.72
2	Norway 1995	98	4199	131	3687	99	1868	2	8.84	1819	3	9.85
3	Iceland 1995	98	3306	137	2585	75	1252	3	8.62	1333	4	9.62
4	Cyprus 1995	93	4360	106	4196	99	2048	3	8.81	2148	4	9.82
5	Greece 1995	95	5547	167	5075	138	2527	3	8.58	2548	4	9.56
6	Scotland 1995	95	5268	130	4924	109	2460	4	8.68	2464	5	9.68
	Secondary school											
7	Norway 2015	99	9342	142	9279	137	4712	8	13.72	4567	9	14.72
8	Norway 1995	96	1295	38	1116	28	547	6	12.86	569	7	13.84
9	Iceland 1995	99	3063	118	2404	67	1219	7	12.63	1185	8	13.64
10	Sweden 1995	96	5753	115	5375	105	3578	7	13.89	1797	8	14.90
11	Cyprus 1995	93	5398	55	5398	55	2646	7	12.71	2752	8	13.72
12	Scotland 1995	95	5286	122	5245	120	2643	Sec. 1	12.68	2602	Sec. 2	13.68
13	Slovak Rep. 1995	93	6586	145	6586	145	3296	7	13.22	3290	8	14.21

The complier samples contain students who were in the formally correct grade. The analysis samples contain those compliers who were at schools with at least 10 students in the lower and 10 students in the upper grade in the samples. The grade labels indicate what the lower and upper grades were called in the respective countries. The mean age is depicted in years

2015, the books at home variable was assessed through student questionnaires and ranged from $1 = $ none or very few (0–10 books) to $5 = $ enough to fill three or more bookcases (more than 200 books). We aggregated the books at home variable while removing missing values. Between 0% (in the Norway 1995 secondary school sample) and 16% (in the Greece 1995 primary school sample) of students in the analyses samples had not responded to the books at home questionnaire item. However, the aggregated number of books variable was available for all schools in all samples, except for two schools in the Scotland 1995 secondary school sample. These two schools were excluded from the analyses that included the socioeconomic school composition.

Statistical analyses

We conducted three sets of analyses. First, we ran regression discontinuity analyses with a student-level focus to illustrate general age and added-year effects on mathematics and science achievement and to ensure that regression discontinuity analyses prerequisites were met. Second, we ran school-level analyses to disentangle the between-school differences in school effectiveness and student selection. These constitute the central analyses in the present study. Third, we regressed this between-school variation in selection and added-year effects on schools' socioeconomic composition. We ran separate analyses for all 13 samples and for both outcome domains. We used Mplus version 8.6 (Muthén & Muthén, 2017) for the analyses and the R package MplusAutomation (Hallquist et al., 2018) to handle the Mplus results. We applied the student sampling weights "TOTWGT" to account for the stratified complex sampling designs (Meinck, 2020) and accommodated the nesting of students in schools in all analyses. The analyses were run separately for the five plausible values and the results were combined using Rubin's (1987) rules.

Regression discontinuity analyses with student-level focus

In the present study, the regression discontinuity approach implies regressing the achievement score Y of student i on the added-year dummy D and age A. To consider the clustering of students in schools, we ran a two-level regression discontinuity model where student i was nested in school j:

$$Y_{ij} = \beta_0 + \beta_1 D_{ij} + \beta_2 A_{ij} + r_{ij}. \tag{1}$$

This regression disentangles the general effect of an added year of schooling β_1 from the general effects of an additional month of age β_2 across schools (see Fig. 1). The intercept β_0 reflects the estimated achievement score of the oldest students in the lower grades ($A = 0$ and $D = 0$). This two-level model estimates the same intercept β_0, schooling effect β_1, and age effect β_2 across schools (i.e., fixed intercept and slopes). A number of prerequisites must be met to estimate these models and interpret them as robust regression discontinuity results (Angrist & Pischke, 2009, 2015; Schochet et al., 2010). These considerations and findings are summarized in Appendix A. From the results presented in Appendix A, we concluded that the regression discontinuity prerequisites were met and that the approach was applicable for all samples.

Regression discontinuity analyses with school-level focus

Since the main aim of the present study is to estimate the between-school variation in added-year and student selection effects within countries, we consequently extended the two-level regression model with random effects for the intercept and added-year slope:

$$Y_{ij} = \beta_{0j} + \beta_{1j}D_{ij} + \beta_2 A_{ij} + r_{ij}. \tag{2}$$

Both the intercept and added-year slope were modelled to vary between schools, as follows:

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{3}$$

and

$$\beta_{1j} = \gamma_{10} + u_{1j}. \tag{4}$$

Inserting (3) and (4) in (2) results in:

$$Y_{ij} = \gamma_{00} + u_{0j} + \gamma_{10}D_{ij} + u_{1j}D_{ij} + \beta_2 A_{ij} + r_{ij}. \tag{5}$$

In these models, the intercept γ_{00} reflects the mean achievement score of the oldest students in the lower grades across schools and u_{0j} reflects its between-school variation. The between-school variation in the intercept reflects all sorts of student selection differences between schools as well as prior school effectiveness differences. Therefore, the intercept can be expected to vary between schools in all countries. The parameter γ_{10} reflects the mean effect of an additional year of schooling across schools and u_{1j} reflects its between-school variation—i.e., the degree to which schools vary in the added-year effect on achievement. The added-year effect can be interpreted as a measure of school effectiveness, and the associated between-school variation can accordingly be regarded as an indicator for the equality of school effectiveness. Ideally, schools should not vary in this school effectiveness measure in education systems. The parameter β_2 reflects the general effect of age on student achievement. It should be noted here that this age effect does not vary between schools in the model. To be able to compare the results of the countries, we z-standardized the achievement plausible values for each country before running these school-level analyses.

Regression discontinuity analyses with school-level focus and socioeconomic school composition predictor

To further illustrate the different meanings of between-school variations in intercepts and added-year effects, we extended the school-level model by regressing both the intercept β_{0j} and the added-year slope β_{1j} on socioeconomic school composition C_j, i.e., the aggregated mean number of books at home, as follows:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}C_j + u_{0j} \tag{6}$$

and

$$\beta_{1j} = \gamma_{10} + \gamma_{11}C_j + u_{1j}. \tag{7}$$

After inserting (6) and (7) in (2), this results in:

$$Y_{ij} = \gamma_{00} + \gamma_{01}C_j + u_{0j} + \gamma_{10}D_{ij} + \gamma_{11}C_jD_{ij} + u_{1j}D_{ij} + \beta_2 A_{ij} + r_{ij}. \tag{8}$$

Therefore, the parameter γ_{01} depicts the degree to which the intercept—i.e., the mean achievement of the oldest students in the lower grades—is associated with socioeconomic school composition. This association should be positive in all countries because privileged students, on average, score higher on achievement tests. The parameter γ_{11} reflects to what extent the mean added-year effect—i.e., school effectiveness measure—relates to the socioeconomic school composition. Ideally, added-year effects should be unrelated to schools' socioeconomic composition. In other words, schools should promote the same learning outcomes, regardless of their socioeconomic composition. Just as we had done for the achievement plausible values, we also z-standardized the aggregated number of books variable before running these school-level analyses.

Results

Descriptive statistics

Table 2 summarizes the descriptive statistics for the 13 samples. The fact that the schools in the samples had different socioeconomic student compositions shows in the intra-class-correlations (ICCs) of the number of books at home variable (see column 1 in Table 2). The lowest ICC was observed in the Norway 1995 primary school sample and the highest in the Scotland 1995 secondary school sample. These books at home ICC values can reflect different selection mechanisms, including residential segregation between social groups. Similarly, columns 2–3 show the ICCs of the mathematics and science achievement scores. Again, the lowest ICCs were observed in the Norway 1995 primary school sample and the highest in the Scotland 1995 secondary school sample. These achievement ICCs can also reflect residential segregation mechanisms as well as differences in the effectiveness of schools. Across the three ICCs, the scores for school segregation by socioeconomic status and achievement were rather low in Norway, Iceland, and Cyprus in comparison to Greece, Scotland, Sweden, and the Slovak Republic.

The mathematics and science score distributions in the lower and upper grades are depicted in columns 4–11 in Table 2. Within domains and countries, the students in the upper grades scored higher than the ones in the lower grades. These differences in means reflect that they were both older and had attended school for an additional year. By tendency, these gaps between younger and older students were higher in the primary than in the secondary school samples. It should be noted here that the secondary school samples did not show systematically higher means than the primary school samples because TIMSS scales achievement scores independently for primary and secondary school populations. Within both grades and domains, the countries differed in their mean achievement scores. These between-sample differences in mean achievement levels reflect all sorts of heterogeneity, including the fact that the samples differed slightly in terms of attended grades and ages (see Table 1).
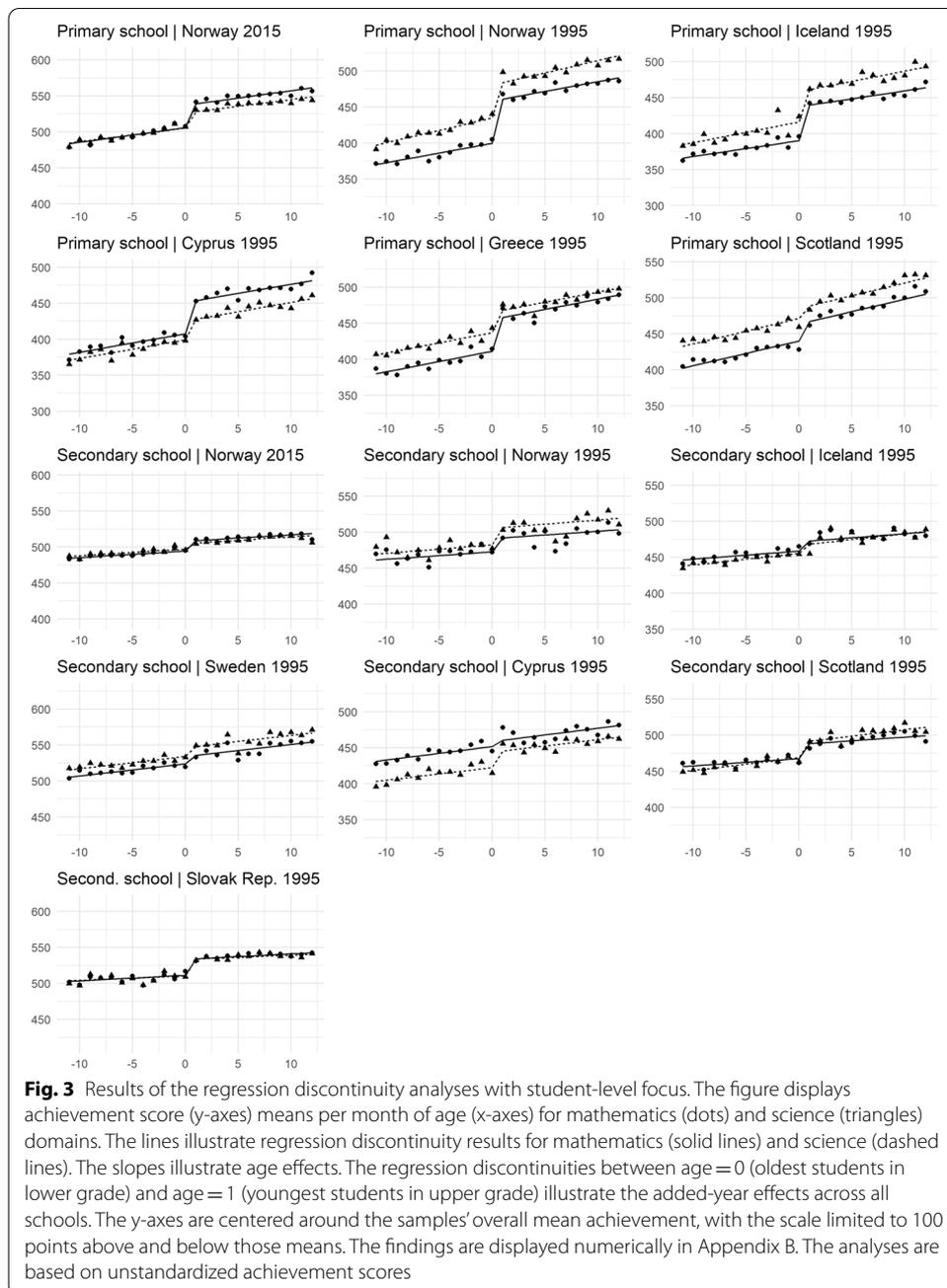
Results of the regression discontinuity analyses with student-level focus

In the first set of analyses, we ran regression discontinuity models that disentangled the age and added-year effects without displaying the between-school differences. Figure 3

Table 2 Descriptive statistics of the 13 samples

School level Country	Books	Mathematics	Science	Lower grade				Upper grade			
				Mathematics		Science		Mathematics		Science	
	ICC	ICC	ICC	M	SD	M	SD	M	SD	M	SD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Primary school											
1 Norway 2015	0.065	0.078	0.079	492.76	71.76	492.85	68.96	548.25	70.49	536.96	62.72
2 Norway 1995	0.021	0.079	0.076	384.48	82.48	417.05	102.70	475.78	73.12	503.48	89.07
3 Iceland 1995	0.029	0.083	0.108	379.74	75.46	402.99	94.09	454.20	75.53	479.26	88.69
4 Cyprus 1995	0.038	0.092	0.102	394.26	91.04	385.72	80.21	466.35	90.91	440.81	80.11
5 Greece 1995	0.087	0.209	0.190	391.21	98.63	418.31	87.24	467.14	97.04	478.15	83.53
6 Scotland 1995	0.153	0.196	0.225	423.27	87.11	454.33	98.67	487.90	90.17	509.75	97.03
Secondary school											
7 Norway 2015	0.088	0.084	0.101	487.45	65.23	489.64	75.73	512.52	69.69	509.46	77.26
8 Norway 1995	0.048	0.081	0.084	466.94	68.75	476.57	75.19	495.81	76.33	511.90	75.92
9 Iceland 1995	0.037	0.080	0.098	454.84	66.67	447.34	73.75	479.29	71.71	475.45	79.08
10 Sweden 1995	0.086	0.200	0.170	515.37	69.53	525.04	74.90	545.24	74.22	558.00	80.25
11 Cyprus 1995	0.026	0.101	0.112	442.36	91.62	413.96	90.37	472.04	91.94	456.11	90.88
12 Scotland 1995	0.172	0.218	0.217	461.79	77.36	458.70	88.74	494.18	80.12	501.94	91.55
13 Slovak Rep. 1995	0.091	0.178	0.202	506.43	72.55	506.61	77.19	537.53	74.11	535.52	78.26

The columns 1–3 contain the between-school intra-class-correlations (ICCs) of the following variables: books at home, mathematics, and science achievement. The columns 4–11 contain the means and standard deviations of the mathematics and science scores in the lower and upper grades. These descriptive analyses used unstandardized achievement scores

Fig. 3 Results of the regression discontinuity analyses with student-level focus. The figure displays achievement score (y-axes) means per month of age (x-axes) for mathematics (dots) and science (triangles) domains. The lines illustrate regression discontinuity results for mathematics (solid lines) and science (dashed lines). The slopes illustrate age effects. The regression discontinuities between age = 0 (oldest students in lower grade) and age = 1 (youngest students in upper grade) illustrate the added-year effects across all schools. The y-axes are centered around the samples' overall mean achievement, with the scale limited to 100 points above and below those means. The findings are displayed numerically in Appendix B. The analyses are based on unstandardized achievement scores

graphically illustrates the results of these analyses for the 13 samples (see Appendix B for numerical results). It should be noted here that the y-axes in the panels in Fig. 3 are centered around the samples' overall means with the scale limits of 100 points above and below these means. Therefore, the y-axes of the panels are comparable even though the samples' general achievement levels differ. The results for mathematics are displayed as solid lines and those for science as dashed lines. The fact that the mean scores by months of age (dots for mathematics and triangles for science) deviated only slightly and unsystematically from the regression lines illustrates that the regression models reflect the data well. As was expected, we observed more deviations in the smaller samples (cf.

Table 1). Figure 3 illustrates that we always found positive effects of age, as indicated by the slopes, and positive effects of an additional year of schooling, as indicated by the regression discontinuities between the oldest students in the lower grades (age = 0) and the youngest students in the upper grades (age = 1). In some cases, we observed very similar achievement levels, age, and added-year effects in the mathematics and science domains, which is reflected in an overlap of regression lines. In other samples, there were greater differences between the two domains.

Conforming to the literature, we observed greater age effects, as indicated by steeper slopes, and greater added-year effects, as indicated by greater regression discontinuities, in the younger samples in comparison to the older ones (see Fig. 3). This can be explained by the fact that the greatest age effects are to be expected when brain maturation proceeds most rapidly—i.e., in young children. However, the fact that the primary and secondary school tests were scaled independently prohibits direct comparisons.

Results of the regression discontinuity analyses with school-level focus

Since the aim of this study is to disentangle the between-school variation in student selection and school effectiveness, we ran two-level regression discontinuity models, as in Eq. (5). The results are summarized in Table 3 for mathematics and in Table 4 for science. In contrast to the previously presented unstandardized results, we used the *z*-standardized achievement scores in these analyses in order to make the results more comparable across the different samples. Hence, intercepts reflect the mean *z*-standardized achievement scores of the oldest students in the lower grade across schools. The variance of the intercepts depicts to what extent schools varied in these achievement levels. This parameter can thus reflect both the between-school differences in student selection and in prior school effectiveness. We found statistically significant between-school variation in intercepts in all samples. We observed lower degrees of between-school variation in intercepts in Norway, Iceland, Sweden, and Cyprus in comparison to Scotland, Greece, and the Slovak Republic (see Tables 3 and 4). We did not observe systematic differences between the primary and secondary school samples from the same countries.

Added-year slope parameters, shown in Tables 3 and 4, reflect the mean effects of an additional year of schooling across schools. At the primary school level, these mean added-year effects ranged between 0.305 standard deviations (SD) in Scotland 1995 and 0.634 SD in Norway 1995 in mathematics and between 0.202 SD in Scotland 1995 and 0.448 SD in Iceland 1995 in science. At the secondary school level, the mean added-year effects ranged between 0.089 SD in Cyprus 1995 and 0.332 SD in the Slovak Republic 1995 in mathematics and between 0.084 SD in Norway 2015 and 0.308 SD in the Slovak Republic 1995 in science. Consistent with the literature and student-level findings in Fig. 2, we observed moderate to large added-year effects at the primary school level and small to moderate added-year effects at the secondary school level.

In our models, the variances of the added-year slopes reflect how much schools varied in added-year effects after taking into account the between-school differences in student selection, i.e., the intercepts. Consequently, these parameters reflect to what extent education systems attained equal school effectiveness at all schools. In both mathematics and science, we observed low degrees of between-school variation in added-year slopes

Table 3 Two-level regression discontinuity results for mathematics achievement as an outcome

Level in model Parameter	Primary school						Secondary school						
	Norway 2015	Norway 1995	Iceland 1995	Cyprus 1995	Greece 1995	Scotland 1995	Norway 2015	Norway 1995	Iceland 1995	Sweden 1995	Cyprus 1995	Scotland 1995	Slovak Rep. 1995
	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
School level													
Intercept γ_{00}	− 0.239* (0.036)	− 0.340* (0.047)	− 0.301* (0.052)	− 0.268* (0.049)	− 0.239* (0.047)	− 0.212* (0.048)	− 0.118* (0.037)	− 0.146 (0.076)	− 0.142* (0.049)	− 0.058 (0.043)	− 0.021 (0.049)	− 0.168* (0.049)	− 0.123* (0.041)
Var. intercept u_{0j}	0.080* (0.015)	0.081* (0.018)	0.073* (0.018)	0.086* ((0.014)	0.236* (0.041)	0.187* (0.040)	0.085* (0.013)	0.043* (0.022)	0.062* (0.016)	0.127* (0.026)	0.093* (0.018)	0.182* (0.033)	0.183* (0.036)
Mean added-year slope $\gamma_{10}D_{ij}$	0.397* (0.049)	0.634* (0.070)	0.576* (0.069)	0.424* (0.071)	0.438* (0.052)	0.305* (0.059)	0.188* (0.051)	0.224 (0.116)	0.191* (0.083)	0.247* (0.085)	0.089 (0.056)	0.296* (0.067)	0.332* (0.056)
Var. added-year slope $u_{1j}D_{ij}$	0.056* (0.014)	0.012 (0.012)	0.009 (0.015)	0.021 (0.013)	0.087* (0.021)	0.056* (0.015)	0.034* (0.011)	0.080 (0.045)	0.102* (0.039)	0.315* (0.045)	0.014 (0.011)	0.184* (0.042)	0.155* (0.028)
Student level													
Mean age slope $\beta_2 A_{ij}$	0.027* (0.003)	0.030* (0.005)	0.024* (0.005)	0.026* (0.005)	0.028* (0.004)	0.033* (0.004)	0.014* (0.003)	0.010 (0.009)	0.016* (0.005)	0.012* (0.004)	0.019* (0.004)	0.006 (0.004)	0.006* (0.003)
Residual variance r_{ij}	0.761* (0.016)	0.659* (0.022)	0.695* (0.030)	0.705* (0.018)	0.600* (0.017)	0.629* (0.017)	0.865* (0.016)	0.826* (0.042)	0.850* (0.034)	0.691* (0.018)	0.842* (0.024)	0.675* (0.019)	0.674* (0.016)

The columns depict the results of the two-level regression model described in Eq. (5) for the different samples. The rows depict the parameters with standard errors (SE) in parentheses. The analyses are based on z-standardized achievement scores

*$p<0.050$

Table 4 Two-level regression discontinuity results for science achievement as an outcome

Level in model Parameter	Primary school						Secondary school						
	Norway 2015	Norway 1995	Iceland 1995	Cyprus 1995	Greece 1995	Scotland 1995	Norway 2015	Norway 1995	Iceland 1995	Sweden 1995	Cyprus 1995	Scotland 1995	Slovak Rep. 1995
	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
School level													
Intercept γ_{00}	−0.181* (0.035)	−0.231* (0.049)	−0.264* (0.058)	−0.196* (0.044)	−0.167* (0.047)	−0.160* (0.055)	−0.064 (0.038)	−0.194* (0.078)	−0.150* (0.054)	−0.056 (0.043)	−0.083 (0.053)	−0.158* (0.049)	−0.118* (0.043)
Var. intercept u_{0j}	0.082* (0.016)	0.082* (0.019)	0.106* (0.026)	0.091* (0.015)	0.207* (0.036)	0.229* (0.042)	0.096* (0.016)	0.059* (0.022)	0.086* (0.020)	0.113* (0.023)	0.099* (0.018)	0.178* (0.029)	0.192* (0.032)
Mean added-year slope $\gamma_{10}D_{ij}$	0.294* (0.045)	0.412* (0.076)	0.448* (0.082)	0.298* (0.072)	0.309* (0.052)	0.202* (0.058)	0.084 (0.052)	0.276* (0.114)	0.157 (0.085)	0.249* (0.078)	0.237* (0.054)	0.275* (0.061)	0.308* (0.053)
Var. added-year slope $u_{1j}D_{ij}$	0.039* (0.011)	0.022 (0.015)	0.011 (0.019)	0.018 (0.011)	0.076* (0.023)	0.034* (0.015)	0.048* (0.015)	0.048 (0.030)	0.104* (0.040)	0.233* (0.037)	0.014 (0.008)	0.101* (0.029)	0.099* (0.021)
Student level													
Mean age slope $\beta_2 A_{ij}$	0.028* (0.003)	0.034* (0.004)	0.026* (0.006)	0.029* (0.005)	0.031* (0.003)	0.031* (0.004)	0.013* (0.004)	0.012 (0.007)	0.021* (0.005)	0.012* (0.004)	0.017* (0.004)	0.013* (0.004)	0.005 (0.003)
Residual variance r_{ij}	0.797* (0.018)	0.744* (0.023)	0.732* (0.029)	0.704* (0.021)	0.614* (0.016)	0.663* (0.020)	0.857* (0.017)	0.836* (0.046)	0.830* (0.039)	0.735* (0.019)	0.805* (0.022)	0.705* (0.021)	0.692* (0.017)

The columns depict the results of the two-level regression model described in Eq. (5) for the different samples. The rows depict the parameters with standard errors (SE) in parentheses. The analyses are based on z-standardized achievement scores

*p < 0.050

in primary school samples, especially in Norway, Iceland, and Cyprus 1995. At the secondary school level, we observed similarly low degrees of between-school variation in added-year slopes in Norway 1995 and 2015 as well as Cyprus 1995 samples. More pronounced between-school variations in these added-year effects were found in the secondary school samples in Sweden, Scotland, and the Slovak Republic 1995.

It is interesting to contrast the two Norway samples for primary grades. Although the mean added-year effect was substantially smaller in 2015 in comparison to 1995, the associated between-school variability increased, becoming statistically significant. In this comparison, however, one should be reminded that Norway participated with second and third graders in 1995 but with fourth and fifth graders in 2015. Similarly, in the interpretation of the results of the secondary school sample in Sweden 1995, it should be considered that this sample contained about twice as many lower than upper grade students. This might point to sampling issues that limit the interpretability of the findings.

Age slope parameters reflect the difference that one month of age makes for achievement scores, independent from grade level and school attended. At the primary school level, these slopes ranged between 0.024 SD in Iceland 1995 and 0.033 SD in Scotland 1995 in mathematics and between 0.026 SD in Iceland 1995 and 0.034 SD in Norway 1995 in science. This means that the effects of one *year* of age (i.e., multiplied by 12) on mathematics and science achievement ranged from moderate to large effect sizes (0.288–0.408 SD). At the secondary school level, age slopes ranged between 0.006 SD in Scotland and the Slovak Republic 1995 and 0.019 SD in Cyprus 1995 in mathematics and between 0.005 SD in the Slovak Republic 1995 and 0.021 SD in Iceland 1995 in science. The effects of one *year* of age on mathematics and science achievement therefore ranged from small to moderate effect sizes (0.060–0.252 SD). Hence, we found more pronounced age effects in younger groups, which conforms to the literature and student-level findings (see Fig. 2).

The residual variances shown in Tables 3 and 4 reflect the degree to which student achievement scores deviated from the values that were predicted on the basis of student age, grade, and attended school.

Results of the regression discontinuity analyses with school-level focus and the socioeconomic school composition predictor

The results of further regression discontinuity analyses, which included the socioeconomic school composition predictor as in Eq. (8), are depicted in Tables 5 and 6 for mathematics and science, respectively. Both socioeconomic school composition and achievement scores were *z*-standardized in these analyses. The results showed positive and, with one exception, significant associations between socioeconomic composition and achievement intercepts of schools. In other words, the more privileged the student composition of a school, the higher the mean achievement of its oldest students in lower grades. These regression coefficients ranged between moderate to large effect sizes in both mathematics (between 0.270 SD in Iceland 1995 secondary school sample and 1.188 SD in Cyprus 1995 secondary school sample) and science (between 0.399 SD in Norway 1995 primary school sample and 1.175 SD in Cyprus 1995 secondary school sample).

Table 5 Two-level regression discontinuity results for mathematics achievement as an outcome and socioeconomic school composition as a predictor

Level in model Parameter	Primary school						Secondary school						
	Norway 2015	Norway 1995	Iceland 1995	Cyprus 1995	Greece 1995	Scotland 1995	Norway 2015	Norway 1995	Iceland 1995	Sweden 1995	Cyprus 1995	Scotland 1995	Slovak Rep. 1995
	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
School level													
Intercept γ_{00}	−0.231* (0.029)	−0.339* (0.046)	−0.295* (0.050)	−0.256* (0.049)	−0.229* (0.045)	−0.192* (0.038)	−0.094* (0.030)	−0.111 (0.072)	−0.122* (0.052)	−0.043 (0.035)	−0.009 (0.038)	−0.151* (0.040)	−0.113* (0.038)
Var. intercept u_{0j}	0.034* (0.008)	0.067* (0.015)	0.053* (0.013)	0.080* (0.014)	0.186* (0.046)	0.084* (0.020)	0.030* (0.006)	0.017 (0.016)	0.055* (0.016)	0.058* (0.012)	0.040* (0.009)	0.076* (0.017)	0.136* (0.033)
Mean added-year slope $\gamma_{10}D_{ij}$	0.399* (0.049)	0.634* (0.070)	0.579* (0.069)	0.421* (0.070)	0.437* (0.052)	0.305* (0.059)	0.186* (0.051)	0.262* (0.120)	0.210* (0.088)	0.253* (0.084)	0.089 (0.056)	0.302* (0.067)	0.333* (0.055)
Var. added-year slope $u_{1j}D_{ij}$	0.049* (0.013)	0.005 (0.010)	0.008 (0.014)	0.020 (0.013)	0.082* (0.020)	0.050* (0.014)	0.028* (0.010)	0.058 (0.032)	0.092* (0.036)	0.293* (0.045)	0.010 (0.009)	0.154* (0.036)	0.138* (0.027)
Intercept on comp $\gamma_{01}C_j$	0.788* (0.087)	0.422* (0.173)	0.599* (0.149)	0.337* (0.145)	0.624* (0.148)	0.755* (0.086)	0.726* (0.065)	0.500* (0.188)	0.270 (0.144)	0.815* (0.106)	1.188* (0.200)	0.707* (0.080)	0.615* (0.098)
Add.-y. slope on comp $\gamma_{11}C_jD_{ij}$	−0.181 (0.108)	0.196 (0.159)	−0.110 (0.158)	−0.066 (0.140)	0.031 (0.099)	0.013 (0.083)	−0.027 (0.087)	0.299 (0.282)	0.296 (0.265)	0.213 (0.209)	0.044 (0.168)	0.192 (0.111)	0.211 (0.108)
Student level													
Mean age slope β_2A_{ij}	0.027* (0.003)	0.030* (0.004)	0.024* (0.005)	0.026* (0.005)	0.028* (0.004)	0.033* (0.004)	0.014* (0.003)	0.009 (0.009)	0.016* (0.005)	0.012* (0.004)	0.019* (0.004)	0.005 (0.004)	0.006* (0.003)
Residual variance r_{ij}	0.762* (0.016)	0.661* (0.021)	0.695* (0.030)	0.705* (0.018)	0.601* (0.016)	0.630* (0.017)	0.866* (0.016)	0.827* (0.041)	0.851* (0.034)	0.691* (0.018)	0.843* (0.025)	0.678* (0.020)	0.675* (0.016)

The columns depict the results of the two-level regression model described in Eq. (8) for the different samples. The rows depict the parameters with standard errors (SE) in parentheses. The analyses are based on z-standardized achievement and school composition scores

*$p < 0.050$

Table 6 Two-level regression discontinuity results for science achievement as an outcome and socioeconomic school composition as a predictor

Level in model Parameter	Primary school						Secondary school						
	Norway 2015	Norway 1995	Iceland 1995	Cyprus 1995	Greece 1995	Scotland 1995	Norway 2015	Norway 1995	Iceland 1995	Sweden 1995	Cyprus 1995	Scotland 1995	Slovak Rep. 1995
	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
School level													
Intercept γ_{00}	− 0.172* (0.026)	− 0.229* (0.049)	− 0.255* (0.053)	− 0.174* (0.042)	− 0.156* (0.043)	− 0.132* (0.039)	− 0.039 (0.030)	− 0.139* (0.069)	− 0.108* (0.054)	− 0.043 (0.035)	− 0.074 (0.041)	− 0.139* (0.036)	− 0.107* (0.039)
Var. intercept u_{0j}	0.030* (0.006)	0.068* (0.017)	0.069* (0.020)	0.073* (0.014)	0.149* (0.036)	0.075* (0.015)	0.031* (0.007)	0.014 (0.012)	0.065* (0.017)	0.048* (0.009)	0.046* (0.012)	0.042* (0.010)	0.133* (0.029)
Mean added-year slope $\gamma_{10}D_{ij}$	0.297* (0.045)	0.411* (0.076)	0.452* (0.082)	0.291* (0.071)	0.308* (0.052)	0.198* (0.058)	0.084 (0.053)	0.287* (0.114)	0.156 (0.092)	0.257* (0.077)	0.238* (0.055)	0.281* (0.061)	0.309* (0.053)
Var. added-year slope $u_{1j}D_{ij}$	0.031* (0.009)	0.012 (0.013)	0.007 (0.016)	0.019 (0.011)	0.069* (0.023)	0.029* (0.013)	0.039* (0.013)	0.042 (0.025)	0.102* (0.041)	0.215* (0.036)	0.011 (0.008)	0.064* (0.021)	0.090* (0.020)
Intercept on comp $\gamma_{01}C_j$	0.841* (0.092)	0.399* (0.184)	0.841* (0.154)	0.598* (0.121)	0.682* (0.139)	0.954* (0.090)	0.778* (0.072)	0.769* (0.252)	0.560* (0.179)	0.800* (0.101)	1.175* (0.202)	0.783* (0.068)	0.715* (0.096)
Add.-y. slope on comp $\gamma_{11}C_jD_{ij}$	− 0.114 (0.100)	0.228 (0.182)	− 0.237 (0.172)	− 0.154 (0.130)	− 0.011 (0.093)	− 0.073 (0.070)	0.002 (0.092)	− 0.101 (0.290)	0.030 (0.239)	0.156 (0.216)	0.067 (0.168)	0.209* (0.095)	0.099 (0.104)
Student level													
Mean age slope β_2A_{ij}	0.028* (0.003)	0.034* (0.004)	0.026* (0.006)	0.030* (0.005)	0.032* (0.003)	0.031* (0.004)	0.013* (0.004)	0.010 (0.007)	0.021* (0.005)	0.012* (0.004)	0.017* (0.004)	0.014* (0.004)	0.005 (0.003)
Residual variance r_{ij}	0.799* (0.018)	0.746* (0.023)	0.733* (0.029)	0.704* (0.021)	0.615* (0.016)	0.664* (0.020)	0.859* (0.017)	0.837* (0.046)	0.830* (0.039)	0.736* (0.019)	0.806* (0.022)	0.710* (0.021)	0.693* (0.017)

The columns depict the results of the two-level regression model described in Eq. (8) for the different samples. The rows depict the parameters with standard errors (SE) in parentheses. The analyses are based on z-standardized achievement and school composition scores

*$p < 0.050$

In contrast, with one exception, the results showed no significant associations between schools' compositions and added-year effects. Only in the Scotland 1995 secondary school sample, a more privileged student composition was significantly associated with more pronounced added-year effects in science. However, it is important to stress that we did not always observe a significant amount of between-school variation in the added-year effects to begin with (see Tables 3 and 4). Therefore, the results of the association between the schools' compositions and added-year effects should not be overinterpreted for the primary school samples in Norway, Iceland, and Cyprus 1995 and for the secondary school samples in Norway and Cyprus 1995 in both mathematics and science.

Discussion

The main aim of the present study is to utilize two-level regression discontinuity analyses to disentangle the between-school variation in school effectiveness (i.e., added-year effects) from the between-school variation in student selection (i.e., intercepts). To achieve this aim, we used overall 13 samples that included both primary and secondary school level data for seven Nordic and six other European countries obtained from TIMSS 1995 and 2015. In all samples, we found considerable between-school differences in intercepts—i.e., achievement levels. This variation can reflect both selection as well as prior effectiveness differences. Conforming to expectations, the intercepts were higher in schools with a more privileged student composition in all samples. Especially in Sweden, Scotland, and the Slovak Republic, we also found between-school variation in added-year effects—i.e., school effectiveness. However, especially Norway and Cyprus attained a low degree of between-school variation in added-year effects—i.e., quite similar added values of an additional year of schooling across schools. Furthermore, added-year effects were not significantly associated with socioeconomic student composition, except for the Scotland 1995 secondary school level sample. Although schools with a privileged socioeconomic composition usually had higher intercepts, they did not always have higher added-year effects. In addition, even though schools differed in student selection, some countries seemed to be able to attain a high degree of equality of school effectiveness. This is the most central implication of our study.

Norway is emphasized in the study because it had implemented the design with two adjacent grades at the primary and secondary school level in both 1995 and 2015. In brief, the analyses revealed that the return of an added year of schooling was substantially lower in 2015, particularly at the primary school level. Additionally, the between-school variability in this added-year effect increased, becoming significant at the primary school level. However, we are cautious not to draw definite conclusions from this finding. Obviously, Norway participated in TIMSS 2015 with higher grade levels and, hence, older students than in 1995. This, in itself, would already lead to lower expected grade effects. Furthermore, Olsen and Björnsson (2018) illustrated that the age effect is not very robust between cycles of international large-scale assessments. Indirectly, this lack of age effect robustness should also affect the robustness of added-year effects over time.

However, this study has a number of additional central implications. It shows that the regression discontinuity approach can be utilized to measure the equality of school effectiveness independently from selection effects. In comparison with longitudinal approaches for measuring school effectiveness, the regression discontinuity approach

has several key advantages (cf. Luyten et al., 2009; Perry, 2017; Singh, 2020). It can be implemented in a cross-sectional assessment of two adjacent grades from the same schools. This makes it more time- and cost-efficient than longitudinal designs, minimizing attrition issues. Another advantage is that it disentangles schooling from pure age effects. Furthermore, this approach can be implemented in a comparably easy manner both internationally and in repeated cycles. We conclude that the suggested measure for the equality of school effectiveness can serve as a feasible and important evaluation criterion when assessing educational inequalities in school effectiveness research.

In comparison with earlier studies that used TIMSS 1995 data, the main contribution of our study is that a less complex and more targeted model was applied in order to strengthen the interpretability of the estimates of the between-school variation in school effectiveness. Additionally, this study was the first to utilize Norway's extension of the TIMSS 2015 design, and it explicitly tested whether the core assumptions of regression discontinuity designs were met in the investigated countries.

Together with the summarized literature (Luyten, 2006; Luyten & Veldkamp, 2011; Luyten et al., 2017; Olsen & Björnsson, 2018; Webbink & Gerritsen, 2013), our study illustrates that even though an additional year of schooling and age usually has positive effects on achievement outcomes, these effects vary considerably between countries, domains, and grade levels as well as over time. This is an interesting finding in light of the common rule of thumb that "learning gains on most national and international tests during one year are equal to between one-quarter and one-third of a standard deviation" (Woessmann, 2016, p. 6; cf. also OECD, 2019a). Given the findings, this rule of thumb seems overgeneralized. Instead, different benchmarks should be developed for different countries, outcome domains, and school career stages.

Limitations

Our study also has some core limitations. First, we focused on only a small number of countries and on two points in time because international large-scale assessments usually assess only one grade per school. However, it would be interesting to compare these findings with findings from other countries, especially lower-income countries (cf. Heyneman & Loxley, 1983).

Second, our study specifically focused on mathematics and science achievement outcomes. The effects of other outcomes, such as less schooling-related cognitive abilities, might differ (Cahan et al., 2008; Cliffordson & Gustafsson, 2008; Kyriakides & Luyten, 2009; Luyten et al., 2017).

Third, the regression discontinuity approach is only applicable in countries that have only one fixed school entry date per year and high proportions of students in formally correct grades. The literature usually applies a 95% complier threshold (e.g., Cliffordson, 2010; Luyten, 2006)—at 93%, we were slightly more liberal in this study. Furthermore, we simply excluded non-compliers, while some authors suggest more complex procedures such as instrumental variable approaches in a fuzzy regression discontinuity framework (e.g., Hahn et al., 2001; Webbink & Gerritsen, 2013).

Fourth, the two-level regression discontinuity approach for identifying the equality of school effectiveness requires, per country, a substantial number of schools that are large enough. In the present study, we also included samples comprising comparably few

schools (see Table 1), and the results obtained should, hence, be interpreted with appropriate caution. We excluded schools with less than 11 students in upper as well as lower grades because the approach is not applicable to very small schools.

Fifth, we found greater age and added-year effects in primary school samples. However, we would like to stress that these are not directly comparable because the achievement tests are not vertically scaled.

Sixth, our approach requires that the sampling mechanisms do not differ between the lower and upper grades of the same schools. We only included samples where the TIMSS technical reporting did not indicate any sampling issues. However, in the secondary school sample in Sweden 1995, there were approximately twice as many students in the lower than in the upper grade samples (see Table 1), which may indicate sampling issues. Therefore, these findings should be interpreted with caution.

Conclusion

We conclude that the two-level regression discontinuity design can be applied if countries assess large enough samples of students from two adjacent grades at the same schools as well as if the share of students in formally correct grades is high enough in these countries. This can be achieved by extending the international designs of large-scale assessment studies with relatively low additional costs and efforts. The resulting data can be utilized to determine the (equality of) the effectiveness of schooling independent of selection and age effects. In addition, this approach makes it possible to investigate what kinds of schools provide greater added-year effects (e.g., Ali & Heck, 2012; Heck & Moriyama, 2010). For instance, the approach can be used to test if schools with a high instructional quality, certain school resources, or high average teacher competencies attain greater added-year effects, than others.

Consequently, we recommend that countries, such as the Nordic ones, assess additional grades through national extensions of large-scale assessments like TIMSS (cf. also Van Damme et al., 2010). When extending the international design in this manner, it would be advisable to assess the same grades over repeated cycles. Unlike in the case of Norway, this would facilitate more direct comparisons over time. Even if countries already have longitudinal data from registries and national assessments, extending international large-scale assessments like this has the advantage of making the findings comparable with those of other countries as well as of allowing rich questionnaire data to be taken into consideration.

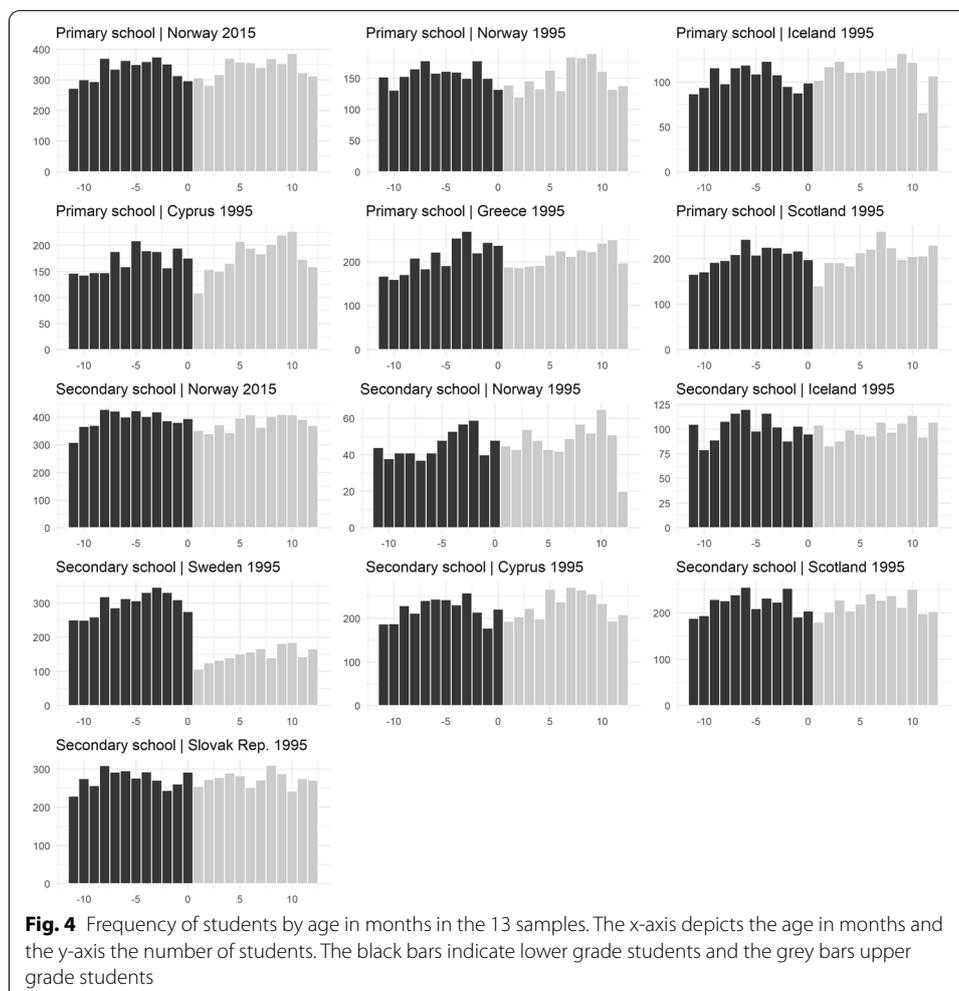## Appendix A: Regression discontinuity prerequisite checks

In order to ensure that the regression discontinuity approach is applicable and delivers robust results, we followed the standards of the What Works Clearinghouse (Schochet et al., 2010) and ran numerous prerequisite check analyses. These analyses, which are described in detail in this section, showed that the regression discontinuity approach could be applied to the current case of achievement discontinuity in lower- and upper-grade students from the same schools as well as to the datasets at hand.

Running variable requirements

The running or forcing variable, in this case age in months, should have at least four unique values below and above the threshold and should not be confounded with other interventions. In the present study, the running variable age consisted of 12 unique values in both lower and upper grades. Furthermore, it seems unlikely that age might be confounded with additional interventions because students of different ages have, within the same grades, the same curricula and teachers.

Integrity of the running variable

It should be ensured that the running variable, in this case age, is not systematically manipulated to change the treatment assignment—i.e., school entry. Manipulating both the birth date of a child and the scoring rule—i.e., school entry regulations—is unlikely. School entry rules are determined by policymakers and rarely changed. The months of age frequency did not systematically change around the threshold (see Fig. 4), which could otherwise be interpreted as an indication of systematic birth date manipulations by parents. Furthermore, the frequency of non-compliers was below 8% in all 13 samples.
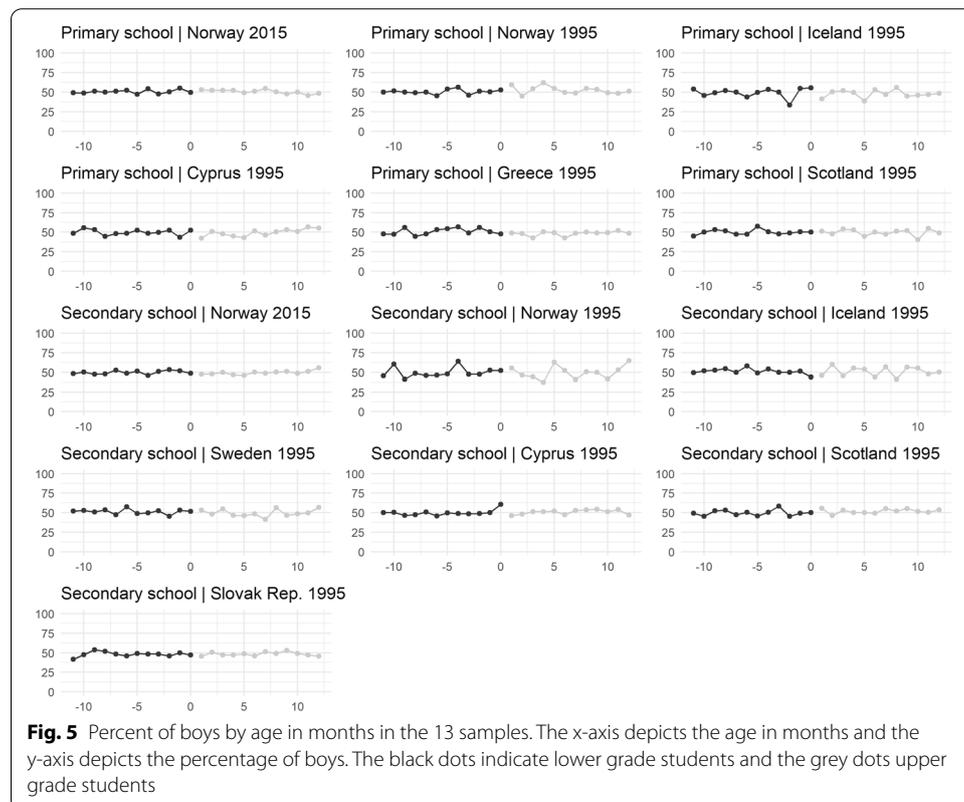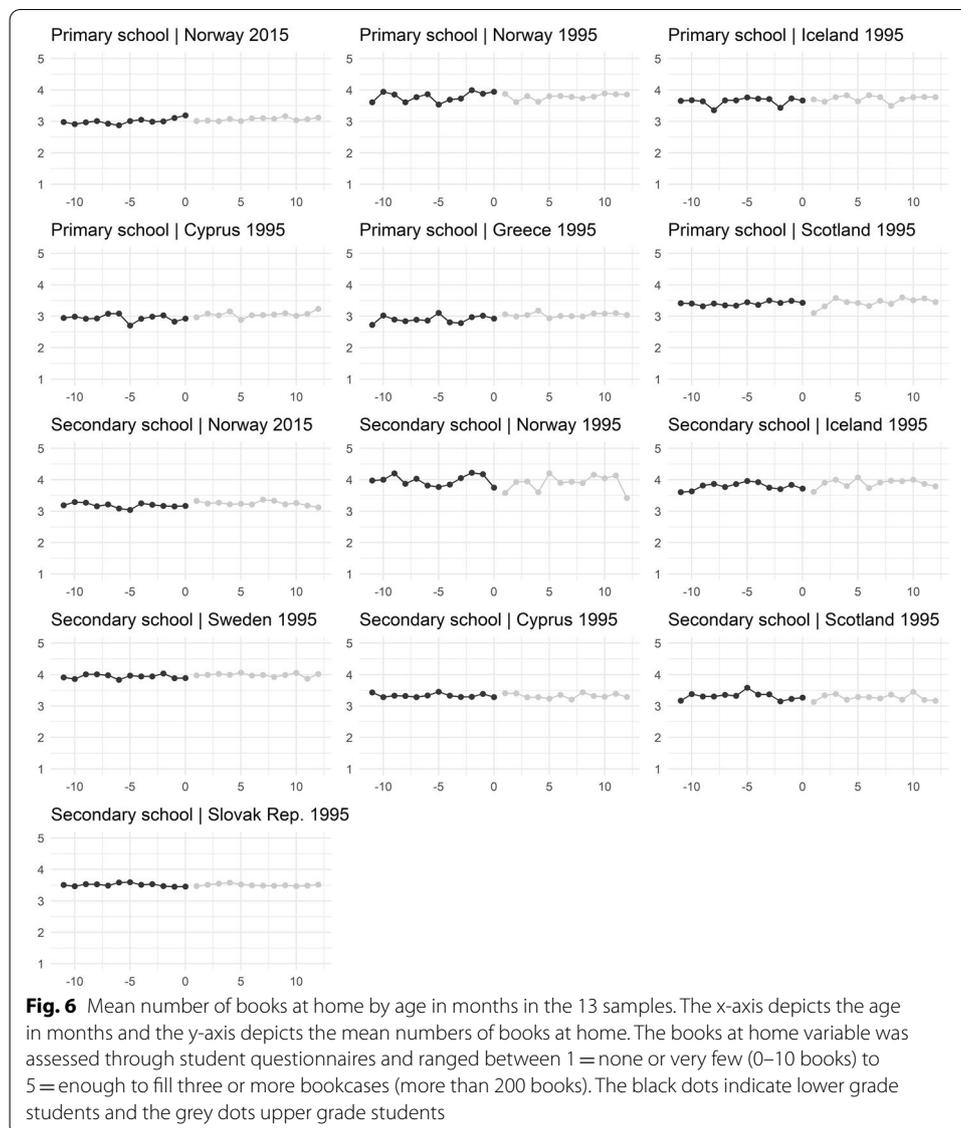

Fig. 4 Frequency of students by age in months in the 13 samples. The x-axis depicts the age in months and the y-axis the number of students. The black bars indicate lower grade students and the grey bars upper grade students

Low attrition rates

Researchers should also assure that attritions by treatment status, running variable, and outcome are low so that the generalizability of the findings and the absence of systematic biases can be assumed. In the present study, there was no attrition by treatment status because all students were either in the lower or upper grades. Attrition by age, the running variable, was negligible ($\leq 2\%$; see Measures section) with one exception. There was also no attrition by outcomes because the achievement plausible values were available for all students. The findings are, however, only generalizable to populations who participated in the TIMSS tests and who were not otherwise excluded (see Data section).

Continuity of the relationship between the running variable and the outcome

Furthermore, it should be established that if the treatment does not take place, then a smooth relationship between the running variable and the outcome would be developed at the threshold. In this study, this means that if the younger and older students entered school at the same time and attended the same grade, we should not observe a discontinuity in the achievement regression. This might not be the case if student selection mechanisms differed between the two grades, for instance. We approximated this by showing that correlates of achievement—namely, gender (see Fig. 5) and the number of books at home (see Fig. 6)—did not systematically change at the threshold. In addition, we observed no other achievement discontinuities within the grades (see mean scores by age in months in Fig. 2).

Fig. 5 Percent of boys by age in months in the 13 samples. The x-axis depicts the age in months and the y-axis depicts the percentage of boys. The black dots indicate lower grade students and the grey dots upper grade students

Fig. 6 Mean number of books at home by age in months in the 13 samples. The x-axis depicts the age in months and the y-axis depicts the mean numbers of books at home. The books at home variable was assessed through student questionnaires and ranged between 1 = none or very few (0–10 books) to 5 = enough to fill three or more bookcases (more than 200 books). The black dots indicate lower grade students and the grey dots upper grade students

Appropriate functional form and bandwidth

Finally, it should be ascertained that the functional form assumption holds empirically and that the effects do not drastically change when investigating subsets of cases closer to the threshold. Our model assumed the same linear effects of age in both grades, after controlling for the added-year effect. We ran alternative models: one with an additional quadratic term for age (see Table 7), one with an additional interaction term between age and grade (see Table 8), and one with a reduced sample of students in the bandwidth of six months before and six months after the threshold (see Table 9). With one exception for each, the regression parameters of the additional quadratic or interaction terms were not statistically different from zero across

Table 7 Results of the regression discontinuity analyses with an additional quadratic term for age

School level Country	Mathematics					Science				
	Intercept	Added-year slope	Age slope	Slope age × age	Residual variance	Intercept	Added-year slope	Age slope	Slope age × age	Residual variance
	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Primary school										
1 Norway 2015	508.290* (2.878)	31.158* (3.837)	2.092* (0.269)	−0.053* (0.021)	4857.783* (146.033)	507.530* (2.723)	20.988* (3.213)	1.939* (0.246)	−0.050* (0.020)	4182.084* (143.544)
2 Norway 1995	400.642* (4.577)	58.802* (6.575)	2.708* (0.429)	−0.023 (0.031)	6071.253* (244.523)	436.815* (5.931)	45.973* (8.070)	3.444* (0.502)	−0.047 (0.045)	9176.954* (379.212)
3 Iceland 1995	390.209* (4.831)	47.265* (6.063)	2.190* (0.463)	0.001 (0.036)	5505.476* (283.353)	416.878* (6.344)	42.239* (8.368)	2.868* (0.617)	−0.017 (0.045)	8290.446* (383.489)
4 Cyprus 1995	407.570* (5.030)	43.391* (7.691)	2.552* (0.536)	0.000 (0.039)	8282.129* (243.883)	399.291* (4.099)	26.304* (6.385)	2.577* (0.487)	−0.003 (0.038)	6330.032* (203.521)
5 Greece 1995	411.149* (5.219)	43.981* (6.361)	2.849* (0.430)	−0.002 (0.034)	8712.287* (384.589)	437.363* (4.707)	28.318* (5.585)	2.822* (0.377)	−0.011 (0.034)	6729.563* (314.634)
6 Scotland 1995	437.305* (4.963)	23.279* (7.108)	3.373* (0.463)	0.066 (0.038)	7670.485* (368.174)	468.266* (6.323)	13.328 (7.611)	3.460* (0.548)	0.077 (0.046)	9413.789* (472.942)
Secondary school										
7 Norway 2015	495.474* (2.762)	13.539* (3.623)	0.930* (0.233)	−0.021 (0.019)	4598.676* (142.043)	498.220* (3.349)	7.486 (4.276)	0.960* (0.316)	−0.021 (0.024)	5937.613* (186.508)
8 Norway 1995	471.091* (6.027)	18.722* (8.870)	0.957 (0.645)	0.035 (0.060)	5256.438* (310.472)	478.411* (7.733)	23.685* (9.811)	0.998 (0.589)	0.082 (0.055)	5687.513* (316.958)
9 Iceland 1995	462.070* (3.644)	12.573 (6.512)	1.219* (0.411)	−0.064 (0.039)	4689.135* (229.396)	455.833* (4.640)	11.288 (7.470)	1.612* (0.484)	−0.007 (0.041)	5624.910* (300.077)
10 Sweden 1995	522.585* (3.124)	10.044 (6.617)	1.641* (0.332)	0.026 (0.027)	5143.392* (219.066)	532.123* (3.462)	12.144 (6.293)	1.652* (0.348)	0.040 (0.026)	5996.782* (212.391)
11 Cyprus 1995	452.769* (4.429)	7.030 (5.612)	1.877* (0.403)	−0.025 (0.029)	8338.689* (257.359)	424.053* (4.948)	21.472* (5.562)	1.765* (0.366)	−0.035 (0.031)	8246.409* (298.051)
12 Scotland 1995	466.525* (4.336)	19.811* (5.542)	0.979* (0.352)	0.021 (0.029)	5978.791* (245.927)	468.949* (4.769)	21.327* (5.840)	1.774* (0.398)	−0.007 (0.034)	7978.777* (278.448)
13 Slovak Rep. 1995	511.624* (3.500)	22.252* (4.693)	0.768* (0.264)	−0.007 (0.021)	5378.109* (212.859)	510.572* (3.678)	22.934* (4.822)	0.597* (0.279)	−0.006 (0.023)	6012.041* (248.947)

The rows depict the results of the regression model described in Eq. (1) with an additional quadratic term for age for the different samples. The columns depict the parameters with standard errors (SE) in parentheses. The analyses are based on unstandardized achievement scores

$*p < 0.050$

Table 8 Results of the regression discontinuity analyses with an additional interaction term

School level Country	Mathematics					Science				
	Intercept	Added-year slope	Age slope	Added-year slope × age	Residual variance	Intercept	Added-year slope	Age slope	Added-year slope × age	Residual variance
	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Primary school										
1 Norway 2015	509.573* (2.983)	31.881* (3.855)	2.722* (0.364)	− 1.372* (0.511)	4854.862* (143.943)	508.599* (2.819)	21.639* (3.198)	2.509* (0.354)	− 1.244* (0.479)	4183.073* (144.082)
2 Norway 1995	401.509* (4.806)	59.191* (6.565)	3.041* (0.618)	− 0.716 (0.783)	6070.330* (244.478)	438.317* (6.252)	46.654* (7.980)	4.074* (0.856)	− 1.353 (1.105)	9174.963* (378.620)
3 Iceland 1995	391.002* (5.044)	47.339* (6.021)	2.329* (0.673)	− 0.269 (0.850)	5507.248* (281.228)	417.660* (6.659)	42.484* (8.350)	3.137* (0.896)	− 0.567 (1.098)	8288.562* (384.414)
4 Cyprus 1995	407.874* (5.241)	43.504* (7.813)	2.612* (0.702)	− 0.122 (0.980)	8283.338* (245.002)	399.424* (4.436)	26.374* (6.335)	2.623* (0.753)	− 0.100 (0.925)	6315.075* (196.952)
5 Greece 1995	411.825* (5.576)	44.208* (6.443)	3.000* (0.588)	− 0.304 (0.876)	8712.894* (384.708)	437.961* (5.127)	28.584* (5.582)	3.019* (0.585)	− 0.414 (0.879)	6723.025* (317.444)
6 Scotland 1995	436.026* (5.370)	22.288* (7.270)	2.642* (0.605)	1.624 (0.977)	7670.168* (367.595)	467.073* (6.905)	12.260 (7.699)	2.668* (0.835)	1.771 (1.171)	9415.367* (472.825)
Secondary school										
7 Norway 2015	495.823* (2.891)	13.815* (3.657)	1.154* (0.314)	− 0.497 (0.476)	4593.922* (139.978)	498.258* (3.565)	7.682 (4.255)	1.123* (0.455)	− 0.374 (0.574)	5934.767* (194.941)
8 Norway 1995	472.019* (6.548)	18.418* (8.946)	0.874 (0.889)	0.271 (1.444)	5257.594* (310.948)	478.320* (8.354)	22.903* (9.705)	0.365 (0.917)	1.439 (1.321)	5687.932* (313.444)
9 Iceland 1995	463.380* (3.823)	13.328* (6.635)	1.942* (0.617)	− 1.575 (0.966)	4689.279* (228.849)	455.570* (5.062)	11.322 (7.562)	1.614* (0.706)	− 0.023 (1.023)	5625.201* (299.921)
10 Sweden 1995	522.341* (3.235)	9.720 (6.776)	1.405* (0.349)	0.537 (0.651)	5140.396* (222.343)	531.671* (3.606)	11.610 (6.433)	1.267* (0.396)	0.873 (0.680)	5996.555* (212.460)
11 Cyprus 1995	453.798* (4.421)	7.508 (5.759)	2.256* (0.481)	− 0.816 (0.684)	8338.047* (257.122)	424.671* (5.059)	21.912* (5.596)	2.144* (0.504)	− 0.830 (0.747)	8245.741* (297.344)
12 Scotland 1995	465.589* (4.568)	19.475* (5.493)	0.647 (0.535)	0.703 (0.720)	5978.627* (246.252)	468.600* (5.105)	21.338* (5.844)	1.762* (0.586)	0.006 (0.848)	7977.686* (278.244)
13 Slovak Rep. 1995	511.745* (3.631)	22.335* (4.689)	0.841* (0.367)	− 0.160 (0.498)	5378.066* (212.867)	510.504* (3.772)	22.978* (4.864)	0.629 (0.363)	− 0.078 (0.558)	6025.692* (245.040)

The rows depict the results of the regression model described in Eq. (1) with an additional interaction term between added-year and age for the different samples. The columns depict the parameters with standard errors (SE) in parentheses. The analyses are based on unstandardized achievement scores

$*p < 0.050$

Table 9 Results of the regression discontinuity analyses with a smaller age bandwidth

School level *Country*	Mathematics				Science			
	Intercept	Added-year slope	Age slope	Residual variance	Intercept	Added-year slope	Age slope	Residual variance
	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Primary school								
1 Norway 2015	509.421* (3.356)	27.251* (5.251)	2.757* (0.692)	4843.287* (180.137)	509.231* (3.091)	15.993* (4.561)	2.783* (0.641)	4159.761* (170.431)
2 Norway 1995	403.621* (4.857)	51.792* (7.725)	3.963* (1.086)	5911.597* (247.729)	436.027* (7.192)	45.753* (11.774)	3.539* (1.371)	9035.392* (447.899)
3 Iceland 1995	391.100* (5.438)	47.463* (7.975)	2.068 (1.167)	5327.020* (283.000)	419.667* (6.754)	37.130* (10.063)	3.899* (1.498)	8349.602* (485.211)
4 Cyprus 1995	407.582* (5.823)	43.605* (10.145)	2.638 (1.379)	8071.871* (314.508)	400.216* (4.356)	22.304* (8.306)	3.382* (1.204)	6231.820* (275.022)
5 Greece 1995	410.562* (5.632)	45.049* (9.064)	2.445* (1.121)	8770.129* (446.473)	435.162* (5.289)	33.488* (8.271)	1.722 (1.015)	6844.656* (382.501)
6 Scotland 1995	436.203* (5.386)	30.013* (9.533)	2.224 (1.289)	7743.013* (449.571)	469.266* (6.733)	16.042 (10.338)	3.037* (1.419)	9742.469* (560.327)
Secondary school								
7 Norway 2015	495.570* (3.142)	11.598* (5.011)	1.192 (0.653)	4653.471* (195.607)	498.375* (3.914)	5.024 (6.274)	1.201 (0.830)	6009.678* (203.420)
8 Norway 1995	468.089* (7.251)	30.708* (12.623)	− 1.540 (1.639)	4981.909* (430.018)	474.322* (8.024)	39.520* (12.710)	− 2.227 (1.630)	5578.207* (431.858)
9 Iceland 1995	461.823* (4.487)	11.524 (8.858)	1.407 (1.207)	4494.301* (279.431)	456.088* (5.343)	11.292 (9.393)	1.676 (1.317)	5537.853* (390.613)
10 Sweden 1995	520.467* (3.608)	18.161* (8.057)	0.364 (0.989)	5265.781* (243.625)	531.356* (4.115)	17.381* (8.010)	0.792 (1.079)	6233.760* (259.991)
11 Cyprus 1995	445.765* (5.049)	21.228* (8.285)	− 0.784 (1.097)	8294.749* (285.595)	418.189* (5.879)	31.396* (8.934)	− 0.060 (1.206)	8232.642* (313.417)
12 Scotland 1995	467.638* (4.293)	17.809* (7.139)	1.043 (0.878)	5946.453* (280.171)	468.693* (5.112)	21.528* (7.816)	1.524 (1.020)	7962.746* (387.270)
13 Slovak Rep. 1995	513.458* (3.800)	16.595* (6.489)	1.821* (0.822)	5237.842* (223.091)	511.464* (3.939)	18.848* (6.544)	1.305 (0.814)	5839.981* (261.382)

The rows depict the results of the regression model described in Eq. (1) with a reduced age bandwidth (six months before and after threshold) for the different samples. The columns depict the parameters with standard errors (SE) in parentheses. The analyses are based on unstandardized achievement scores

*$p < 0.050$

samples and domains. When only analyzing the subset of students who were closest to the threshold, both age and grade effects were either statistically significant and positive or not significantly different from zero.

Appendix B: Results of the regression discontinuity analyses with student-level focus

See Table 10.

Table 10 Results of the regression discontinuity analyses with student-level focus

School level Country	Mathematics				Science			
	Intercept	Added-year slope	Age slope	Residual variance	Intercept	Added-year slope	Age slope	Residual variance
	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)	Par (SE)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Primary school								
1 Norway 2015	505.825* (2.799)	31.155* (3.834)	2.027* (0.268)	4847.435* (152.463)	505.192* (2.701)	21.000* (3.208)	1.877* (0.244)	4196.172* (148.098)
2 Norway 1995	399.594* (4.435)	58.656* (6.582)	2.694* (0.426)	6074.167* (243.250)	434.680* (5.562)	45.678* (8.137)	3.415* (0.491)	9179.148* (381.007)
3 Iceland 1995	390.246* (4.416)	47.254* (6.041)	2.191* (0.453)	5508.710* (283.667)	416.062* (6.164)	42.318* (8.359)	2.846* (0.604)	8296.467* (384.427)
4 Cyprus 1995	407.561* (5.095)	43.398* (7.674)	2.551* (0.536)	8308.468* (234.948)	399.170* (3.692)	26.277* (6.417)	2.574* (0.470)	6331.527* (193.516)
5 Greece 1995	411.061* (5.081)	43.955* (6.353)	2.846* (0.432)	8712.216* (384.525)	436.896* (4.467)	28.284* (5.590)	2.807* (0.374)	6748.209* (320.059)
6 Scotland 1995	440.219* (4.481)	23.661* (7.087)	3.437* (0.469)	7678.334* (369.882)	471.654* (5.573)	13.743 (7.603)	3.535* (0.542)	9429.816* (475.866)
Secondary school								
7 Norway 2015	494.471* (2.705)	13.539* (3.623)	0.903* (0.235)	4604.117* (145.718)	497.232* (3.067)	7.494 (4.274)	0.932* (0.310)	5938.218* (187.436)
8 Norway 1995	472.801* (5.859)	18.423* (8.902)	1.023 (0.664)	5259.227* (311.208)	482.441* (6.814)	23.011* (9.679)	1.151* (0.584)	5700.460* (316.976)
9 Iceland 1995	458.867* (3.667)	12.836* (6.520)	1.120* (0.408)	4698.816* (229.414)	455.500* (4.507)	11.321 (7.493)	1.602* (0.484)	5626.104* (300.330)
10 Sweden 1995	523.816* (3.089)	10.024 (6.622)	1.684* (0.350)	5142.924* (218.476)	534.063* (3.317)	12.116 (6.298)	1.719* (0.361)	5999.333* (213.270)
11 Cyprus 1995	451.626* (4.651)	6.940 (5.568)	1.856* (0.407)	8341.627* (259.260)	422.464* (5.040)	21.324* (5.530)	1.737* (0.367)	8243.619* (300.971)
12 Scotland 1995	467.560* (4.045)	19.749* (5.527)	1.009* (0.346)	5981.705* (247.161)	468.616* (4.552)	21.346* (5.845)	1.764* (0.396)	7977.925* (278.232)
13 Slovak Rep. 1995	511.306* (3.401)	22.259* (4.688)	0.761* (0.262)	5368.146* (209.076)	510.292* (3.658)	22.936* (4.825)	0.591* (0.282)	6012.676* (248.595)

The rows depict the results of the regression model described in Eq. (1) for the different samples. These results are illustrated graphically in Fig. 2. The columns depict the parameters with standard errors (SE) in parentheses. The analyses are based on unstandardized achievement scores

*$p < 0.050$

the different parts of the written manuscript. Both authors read and approved the final manuscript. Both authors are accountable for the accuracy and integrity of the study.

Authors' information
Isa Steinmann is postdoctoral researcher at the Centre for Educational Measurement (CEMO) at the University of Oslo. Rolf Vegar Olsen is professor at and co-director of CEMO at the University of Oslo.

Availability of data and materials
The TIMSS 1995 datasets analyzed during the current study are available in the repository of the International Association for the Evaluation of Educational Achievement (IEA): https://timssandpirls.bc.edu/timss1995i/Database.html. The Norwegian extension data from TIMSS 2015 are available from the Department of Teacher Education and School Research (ILS) at the University of Oslo on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable since the study used anonymized data that was collected by the International Association for the Evaluation of Educational Achievement (IEA). The IEA assured that all guidelines regarding ethics and data protection were followed.

Consent for publication
The authors consent to the publication of the manuscript in *Large-scale Assessments in Education.*

Competing interests
The authors declare that they have no competing interests.

References

Ali, E., & Heck, R. H. (2012). Comparing the contexts of middle-grade schools, their instructional practices, and their outcomes: A regression discontinuity approach. *NASSP Bulletin, 96*(2), 93–118. https://doi.org/10.1177/0192636512444715

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics*. Princeton University Press.

Angrist, J. D., & Pischke, J.-S. (2015). *Mastering 'metrics*. Princeton University Press.

Baltes, P. B., & Reinert, G. (1969). Cohort effects in cognitive development of children as revealed by cross-sectional sequences. *Developmental Psychology, 1*(2), 169–177. https://doi.org/10.1037/h0026997

Blossfeld, H.-P., Kulic, N., Skopek, J., Triventi, M., Kilpi-Jakonen, E., Vono de Vilhena, D., & Buchholz, S. (2019). Conditions and consequences of unequal educational opportunities in the life course: Results from the cross-national comparative eduLIFE Project. *KZfSS Kölner Zeitschrift Für Soziologie Und Sozialpsychologie, 71*(S1), 399–428. https://doi.org/10.1007/s11577-019-00595-w

Blossing, U., Imsen, G., & Moos, L. (Eds.). (2014). *The Nordic education model*. Springer Netherlands. https://doi.org/10.1007/978-94-007-7125-3

Bosker, R. J., & Scheerens, J. (1994). Alternative models of school effectiveness put to the test. *International Journal of Educational Research, 21*(2), 159–180. https://doi.org/10.1016/0883-0355(94)90030-2

Cahan, S., & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child Development, 60*(5), 1239. https://doi.org/10.2307/1130797

Cahan, S., & Davis, D. (1987). A between-grade-levels approach to the investigation of the absolute effects of schooling on achievement. *American Educational Research Journal, 24*(1), 1–12. https://doi.org/10.3102/00028312024001001

Cahan, S., Greenbaum, C., Artman, L., Deluya, N., & Gappel-Gilon, Y. (2008). The differential effects of age and first grade schooling on the development of infralogical and logico-mathematical concrete operations. *Cognitive Development, 23*(2), 258–277. https://doi.org/10.1016/j.cogdev.2008.01.004

Carlsson, M., Dahl, G. B., & Rooth, D.-O. (2012). *The effect of schooling on cognitive skills* (NBER Working Paper No. 18484). National Bureau of Economic Research. http://www.nber.org/papers/w18484

Cascio, E. U., & Lewis, E. G. (2006). Schooling and the armed forces qualifying test. *Journal of Human Resources, XLI*(2), 294–318. https://doi.org/10.3368/jhr.XLI.2.294

Chapman, C., Armstrong, P., Harris, A., Muijs, D., Reynolds, D., & Sammons, P. (2012). *School effectiveness and improvement research, policy, and practice*. Routledge.

Cliffordson, C. (2010). Methodological issues in investigations of the relative effects of schooling and age on school performance: The between-grade regression discontinuity design applied to Swedish TIMSS 1995 data. *Educational Research and Evaluation, 16*(1), 39–52. https://doi.org/10.1080/13803611003694391

Cliffordson, C., & Gustafsson, J.-E. (2008). Effects of age and schooling on intellectual performance: Estimates obtained from analysis of continuous variation in age and length of schooling. *Intelligence, 36*(2), 143–152. https://doi.org/10.1016/j.intell.2007.03.006

Crone, D. A., & Whitehurst, G. J. (1999). Age and schooling effects on emergent literacy and early reading skills. *Journal of Educational Psychology, 91*(4), 604–614. https://doi.org/10.1037/0022-0663.91.4.604

Figlio, D., Karbownik, K., & Salvanes, K. G. (2016). Education research and administrative data. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 5, pp. 75–138). Elsevier. https://doi.org/10.1016/B978-0-444-63459-7.00002-6

Frønes, T. S., Pettersen, A., Radišić, J., & Buchholtz, N. (2020). Equity, equality and diversity in the Nordic model of education—Contributions from large-scale studies. In T. S. Frønes, A. Pettersen, J. Radišić, & N. Buchholtz (Eds.), *Equity, equality and diversity in the Nordic model of education* (pp. 1–10). Springer International Publishing. https://doi.org/10.1007/978-3-030-61648-9_1

Gormley, W. T., & Gayer, T. (2005). Promoting school readiness in Oklahoma. *Journal of Human Resources, XL*(3), 533–558. https://doi.org/10.3368/jhr.XL.3.533

Hahn, J., Todd, P., & Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica, 69*(1), 201–209. https://doi.org/10.1111/1468-0262.00183

Hallquist, M., Wiley, J., & van Lissa, C. (2018). *MplusAutomation*. https://github.com/michaelhallquist/MplusAutomation

Heck, R. H., & Moriyama, K. (2010). Examining relationships among elementary schools' contexts, leadership, instructional practices, and added-year outcomes: A regression discontinuity approach. *School Effectiveness and School Improvement, 21*(4), 377–408. https://doi.org/10.1080/09243453.2010.500097

Heyneman, S. P., & Loxley, W. A. (1983). The effect of primary-school quality on academic achievement across twenty-nine high- and low-income countries. *American Journal of Sociology, 88*(6), 1162–1194. https://doi.org/10.1086/227799

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics, 142*(2), 615–635. https://doi.org/10.1016/j.jeconom.2007.05.001

Jabr, D., & Cahan, S. (2014). Schooling effects on cognitive development in a difficult environment: The case of refugee camps in the West Bank. *International Studies in Sociology of Education, 24*(2), 165–188. https://doi.org/10.1080/09620214.2014.900949

Jabr, D., & Cahan, S. (2015). Between-context variability of the effect of schooling on cognitive development: Evidence from the Middle East. *School Effectiveness and School Improvement, 26*(3), 441–466. https://doi.org/10.1080/09243453.2014.944546

Jackson, M. V. (Ed.). (2013). *Determined to succeed? Performance versus choice in educational attainment*. Stanford University Press.

Jerrim, J., & Macmillan, L. (2015). Income inequality, intergenerational mobility, and the Great Gatsby Curve: Is education the key? *Social Forces, 94*(2), 505–533. https://doi.org/10.1093/sf/sov075

Kyriakides, L., & Luyten, H. (2009). The contribution of schooling to the cognitive development of secondary education students in Cyprus: An application of regression discontinuity with multiple cut-off points. *School Effectiveness and School Improvement, 20*(2), 167–186. https://doi.org/10.1080/09243450902883870

Luyten, H. (2006). An empirical assessment of the absolute effect of schooling: Regression-discontinuity applied to TIMSS-95. *Oxford Review of Education, 32*(3), 397–429. https://doi.org/10.1080/03054980600776589

Luyten, H., Merrell, C., & Tymms, P. (2017). The contribution of schooling to learning gains of pupils in Years 1 to 6. *School Effectiveness and School Improvement, 28*(3), 374–405. https://doi.org/10.1080/09243453.2017.1297312

Luyten, H., Merrell, C., & Tymms, P. (2020). Absolute effects of schooling as a reference for the interpretation of educational intervention effects. *Studies in Educational Evaluation, 67*, 100939. https://doi.org/10.1016/j.stueduc.2020.100939

Luyten, H., Peschar, J., & Coe, R. (2008). Effects of schooling on reading performance, reading engagement, and reading activities of 15-year-olds in England. *American Educational Research Journal, 45*(2), 319–342. https://doi.org/10.3102/0002831207313345

Luyten, H., Tymms, P., & Jones, P. (2009). Assessing school effects without controlling for prior achievement? *School Effectiveness and School Improvement, 20*(2), 145–165. https://doi.org/10.1080/09243450902879779

Luyten, H., & Veldkamp, B. (2011). Assessing effects of schooling with cross-sectional data: Between-grades differences addressed as a selection-bias problem. *Journal of Research on Educational Effectiveness, 4*(3), 264–288. https://doi.org/10.1080/19345747.2010.519825

Marks, G. N. (2005). Cross-national differences and accounting for social class inequalities in education. *International Sociology, 20*(4), 483–505. https://doi.org/10.1177/0268580905058328

Martin, M. O., & Kelly, D. L. (Eds.). (1996). *Third international mathematics and science study technical report, volume I*. Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Martin, M. O., & Kelly, D. L. (1997). *Third international mathematics and science study technical report, volume II*. Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Martin, M. O., Mullis, I. V. S., & Hooper, Martin. (2016). *Methods and Procedures in TIMSS 2015*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA).

Meinck, S. (2020). Sampling, weighting, and variance estimation. In H. Wagemaker (Ed.), *Reliability and validity of international large-scale assessment* (Vol. 10, pp. 113–129). Springer International Publishing. https://doi.org/10.1007/978-3-030-53081-5_7

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 international results in reading*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA). http://timssandpirls.bc.edu/pirls2016/international-results/

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA). https://timssandpirls.bc.edu/timss2019/international-results/

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.

OECD. (2012). Learning beyond fifteen: Ten years after PISA. *OECD*. https://doi.org/10.1787/9789264172104-en

OECD. (2019a). PISA 2018 results (volume I): What students know and can do. *OECD Publishing*. https://doi.org/10.1787/5f07c754-en

OECD. (2019b). PISA 2018 results (volume II): Where all students can succeed. *OECD Publishing*. https://doi.org/10.1787/b5fd1b8f-en

Olsen, R. V., & Björnsson, J. K. (2018). Fødselsmåned og skoleprestasjoner. In J. K. Björnsson & R. V. Olsen (Eds.), *Tjue år med TIMSS og PISA i Norge. Trender og nye analyser* (pp. 76–93). Scandinavian University Press.

Perry, T. (2017). Inter-method reliability of school effectiveness measures: A comparison of value-added and regression discontinuity estimates. *School Effectiveness and School Improvement, 28*(1), 22–38. https://doi.org/10.1080/09243453.2016.1203799

Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rost, J., Schiefele, U., & Deutsches PISA-Konsortium (Eds.). (2006). *PISA 2003: Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* [PISA 2003: Studies on competence development over the course of a school year]. Waxmann.

Ræder, H. G., Olsen, R. V., & Blömeke, S. (2020). Large-scale assessments in the Norwegian context. In H. Harju-Luukkainen, N. McElvany, & J. Stang (Eds.), *Monitoring student achievement in the 21st century* (pp. 195–206). Springer International Publishing. https://doi.org/10.1007/978-3-030-38969-7_16

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley.

Schochet, P., Cook, T., Deke, J., Imbens, G., Lockwood, J. R., Porter, J., & Smith, J. (2010). *Standards for regression discontinuity designs*. What Works Clearinghouse. https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_rd.pdf

Schütz, G., Ursprung, H. W., & Wößmann, L. (2008). Education policy and equality of opportunity. *Kyklos, 61*(2), 279–308. https://doi.org/10.1111/j.1467-6435.2008.00402.x

Singh, A. (2020). Learning more with every year: School year productivity and international learning divergence. *Journal of the European Economic Association, 18*(4), 1770–1813. https://doi.org/10.1093/jeea/jvz033

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417–453. https://doi.org/10.3102/00346543075003417

Stelzl, I., Merz, F., Ehlers, T., & Remer, H. (1995). The effect of schooling on the development of fluid and cristallized intelligence: A quasi-experimental study. *Intelligence, 21*(3), 279–296. https://doi.org/10.1016/0160-2896(95)90018-7

Strietholt, R., Gustafsson, J.-E., Hogrebe, N., Rolfe, V., Rosén, M., Steinmann, I., & Hansen, K. Y. (2019). The impact of education policies on socioeconomic inequality in student achievement: A review of comparative studies. In L. Volante, S. V. Schnepf, J. Jerrim, & D. A. Klinger (Eds.), *Socioeconomic inequality and student outcomes* (Vol. 4, pp. 17–38). Springer Singapore. https://doi.org/10.1007/978-981-13-9863-6_2

Townsend, T. (2007). *International handbook of school effectiveness and improvement* (Vol. 17). Springer Netherlands.

Van Damme, J., Liu, H., Vanhee, L., & Pustjens, H. (2010). Longitudinal studies at the country level as a new approach to educational effectiveness: Explaining change in reading achievement (PIRLS) by change in age, socio-economic status and class size. *Effective Education, 2*(1), 53–84. https://doi.org/10.1080/19415531003616888

Van de Vijver, F. J. R., & Brouwers, S. A. (2009). Schooling and basic aspects of intelligence: A natural quasi-experiment in Malawi. *Journal of Applied Developmental Psychology, 30*(2), 67–74. https://doi.org/10.1016/j.appdev.2008.10.010

Webbink, H. D., & Gerritsen, S. (2013). *How much do children learn in school?* (Vol. 255). CPB Netherlands Bureau for Economic Policy Analysis.

Woessmann, L. (2016). The importance of school systems: Evidence from international differences in student achievement. *Journal of Economic Perspectives, 30*(3), 3–32. https://doi.org/10.1257/jep.30.3.3

Yang Hansen, K., Gustafsson, J.-E., Rosén, M., Sulkunen, S., Nissinen, K., Kupari, P., Ólafsson, R. F., Björnsson, J. K., Grønmo, L. S., Rønberg, L., & Mejding, J. (2014). Northern lights on TIMSS and PIRLS 2011. *Nordic Council of Ministers*. https://doi.org/10.6027/TN2014-528

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.