Oxford Review of Education
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/core20

An empirical assessment of the absolute effect of schooling: regression-discontinuity applied to TIMSS-95

Hans Luyten [a]
[a] University of Twente , The Netherlands
Published online: 16 Aug 2006.

Oxford Review of Education
Vol. 32, No. 3, July 2006, pp. 397–429

An empirical assessment of the absolute effect of schooling: regression-discontinuity applied to TIMSS-95

Hans Luyten*
University of Twente, The Netherlands

This article provides an illustration of how multilevel modelling, which is the standard method for assessing the 'school effect' in school effectiveness research, can be combined with the regression-discontinuity approach. While multilevel modelling yields only estimates of relative school effects (differences between schools), the regression-discontinuity approach allows the assessment of the absolute effect of schooling (schooling versus no schooling). A combination of both approaches yields an estimate of this absolute effect and its variation between schools. This study applied the method in a secondary analysis of data derived from the Third International Mathematics and Science Study (TIMSS).

Introduction

One of the main findings of the famous 'Coleman Report' (Coleman *et al.*, 1966) was that approximately 90% of all variance in student achievement in the United States of America was due to variation between students within schools, whereas differences between schools could account for only 10%. At the time, this result was considered disappointing, as it seemed to imply that schools had little impact on achievement. Even though many authors (e.g. Brookover *et al.*, 1979; Edmonds, 1979; Mortimore *et al.*, 1988) have been eager to refute this interpretation, the finding that school differences can explain only a small percentage of the variance in student achievement has been repeatedly corroborated in numerous studies (Scheerens & Bosker, 1997). Since the publication of the Coleman Report, the term 'school effect' and the percentage of variance in achievement scores situated at the school level have become synonymous in the research literature on school effectiveness. Authors like Goldstein

*Faculty of Behavioural Sciences, Department of Educational Organisation and Management, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands. Email: j.w.luyten@gw.utwente.nl

ISSN 0305-4985 (print)/ISSN 1465-3915 (online)/06/030397–33

DOI: 10.1080/03054980600776589

(1997) and Coe and Fitz-Gibbon (1998) have pointed out that the term 'school effect' is somewhat misleading, as it expresses only *relative* differences between schools. The relatively small amount of variation between schools does not necessarily imply that education makes only a minor contribution to the development of students. The small amount of variation may be due to the presence of high-quality instruction in most schools. Although little is known about the *absolute* effect of schooling, available findings suggest that the effects are substantial (e.g. Heyns, 1978; Cooper *et al.*, 1996; Tymms *et al.*, 1997; Lindahl, 2001).

The regression-discontinuity approach

Assessing the contribution that education makes to the development of students is complicated by the fact that nearly everyone attends school. Estimating the absolute effect of education should ideally involve comparisons between those who attend school and a control group of people that do not attend school. In practice, this is not feasible. Nevertheless, this article reports estimates of the absolute effect of schooling based on empirical data. These estimates were obtained by applying the regression-discontinuity approach, sometimes referred to as a 'cutting-point design' (Rossi *et al.*, 2004, p. 289). This approach is very useful, provided the precise criteria for assigning persons to the 'treatment' and the 'control' groups are known. Taking the impact of these criteria into account in the data-analysis makes it possible to assess the impact of the treatment most accurately (Trochim, 1984; Shadish *et al.*, 2002; Rossi *et al.*, 2004).

In many countries, admission to school depends primarily on a student's date of birth. In English primary education, for example, students born before September 1 are usually placed in a higher grade than are students born after that date. If data from at least two consecutive grades are available (e.g. achievement scores that relate to a common scale), including the students' dates of birth, it is possible to assess the absolute contribution of schooling to the development of those students. For each grade, the effect of age on achievement (or some other outcome measure) is estimated. A discontinuity is expected between the oldest students in the lower grade and the youngest students in the higher grade. This discontinuity reflects the effect of having received an extra year of schooling (i.e. being in the higher grade). This interpretation is warranted, as the analysis takes into account the impact of the criterion used for assigning students to the higher or the lower grade. As with a randomised experiment, knowledge of the selection process allows relatively simple statistics to yield unbiased effect estimates in the regression-discontinuity approach (Shadish *et al.*, 2002; p. 224). Note, however, that correct modelling of the relation between date of birth and achievement is crucial. If a linear function is estimated, while, in reality, the function is quadratic or cubic the model will be misspecified, and the regression-discontinuity may be biased. It is logical to expect the slope of the relationship between age and achievement to become less steep in the higher grades. This question is addressed in the analyses presented in this article. Findings on the (curvi)linearity of the relationship between date of birth and achievement are presented in the appendix. The regression-discontinuity approach is also based on the assumption that the average level of achievement is similar across cohorts. In this study, two cohorts of students are compared: those in the lower grade and those in the upper grade. It is conceivable that the findings are biased, as the level of achievement differs between cohorts. This issue cannot be resolved with the data available for this study and needs to be addressed in future research.

An additional advantage of the regression-discontinuity approach, and one that is of considerable practical importance, is that—in principle—it does not require controlling for prior achievement, or any other background characteristics, in order to assess the effect of schooling. As in randomised experiments, cross-sectional data should suffice for strong causal inferences. Standard multilevel analyses in school effectiveness research focus on differences in student achievement between schools. Controlling for prior achievement is then essential, in order to adjust for bias resulting from intake differences between schools. When assignment to the treatment or control group is random, however, both groups differ only by chance. Any observed difference on the outcome variable must be due either to the intervention or to random fluctuations. In the regression-discontinuity approach, the exact criterion that determines assignment to treatment or control group is known, and its effect is taken into account in the analysis. Any remaining difference should therefore be due either to the intervention or to random fluctuation. Taking the impact of background variables into account in the analysis nonetheless makes it possible to control for some of the random differences between the treatment and control groups. Furthermore, it is conceivable that the cut-off point coincides with other relevant factors, although, generally speaking, this is unlikely. The main strength of the regression-discontinuity approach is that it predicts an effect at a very specific point on the age continuum. Even so, some alternative explanations cannot be ruled out completely. For example, in England, children who are born in August may be treated differently from children who are born in September, with regard to pre-school provision. In this case, prior achievement adjustments should still be made. Possible peer group effects raise another point. Being the youngest or oldest in a class may have an effect on achievement *per se* (Blatchford *et al.*, 2002). Unfortunately, these questions could not be addressed in the present study, as the data sets analysed do not provide sufficient information. The extent to which ignoring these factors introduces bias into the effect estimates remains a question for future research. The findings presented in this article reveal the amount of difference in achievement between children born in the month before the cut-off point and those born the month thereafter.

Examining the impact of background variables may be worthwhile for other purposes as well. One interesting research question, for example, is the extent to which the effect of schooling varies between students from different backgrounds. Note, however, that analysing whether an effect differs across certain groups of students is not the same as controlling for differences in intake between the treatment and control groups.

To the author's knowledge, the regression-discontinuity approach has thus far been applied only once to estimate the absolute effect of schooling. This application was part of a study by Cahan and Davis (1987), which relates to primary education in Israel, based on data collected in 1973. Its main conclusion is that approximately two thirds of the difference in achievement between grade levels is due to additional schooling, and only one third can be explained by the age difference between the two grades. This finding contrasts sharply with the outcomes of mainstream school effectiveness research, which typically identifies school effects of about 10%, or even less (Scheerens & Bosker, 1997, p. 79). Currently available multilevel software (Snijders & Bosker, 1999) allows the estimation of school-specific discontinuities. A multilevel analysis can estimate the main effect of one extra year of schooling in the same way as did the ordinary regression analysis conducted by Cahan and Davis. In addition, however, multilevel analysis can estimate the variance of this effect across schools. In other words, it estimates the absolute effect of schooling, as well as the extent to which this effect differs from school to school. The following section presents basic mathematical details regarding the application of the regression-discontinuity approach within a multilevel framework.

Modelling the absolute effect of schooling

Equation (1) presents the basic regression-discontinuity model. The coefficients β_1 and β_2 express the effect of age and the effect of one extra year of schooling. It applies to a dataset with students from two consecutive grades. The effect of age is assumed linear and identical in both grades (i.e. no interaction between age and grade).

$$Y_i = \beta_0 + \beta_1(x_i - x_0) + \beta_2 z_i + R_i \tag{1}$$

Where:

Y_i = *outcome measure (e.g. mathematics achievement score for student i)*
x_i = *date of birth, student i*
x_0 = *cut-off value (e.g. September 1985)*
z_i = *grade level, student i (0 if lower grade; 1 if higher grade)*
β_0 = *parameter for comparison group intercept at cut-off*
β_1 = *date of birth effect*
β_2 = *effect of being in the higher grade (i.e. having received an extra year of schooling)*
R_i = *random residual*

When combining the regression-discontinuity approach with multilevel analysis, the intercept (β_0) and the effect of one extra year of schooling (β_2) are allowed to vary across schools:

$$Y_{ij} = \beta_{0j} + \beta_1(x_{ij} - x_0) + \beta_{2j} z_{ij} + R_{ij} \tag{2}$$

j = *index for schools*
i = *index for students within schools*

The intercept and the effect of schooling are now school-dependent. These school-dependent coefficients can be separated into an average coefficient and the school-dependent deviation:

$$\beta_{0j} = \gamma_{00} + U_{0j}$$
$$\beta_{2j} = \gamma_{20} + U_{2j}$$

Substitution leads to the following model:

$$Y_{ij} = \gamma_{00} + \beta_{10}(x_{ij} - x_0) + \gamma_{20} z_{ij} + U_{0j} + U_{2j} z_{ij} + R_{ij} \quad (3)$$

In this equation, γ_{20} expresses the general effect of having received an extra year of schooling, while U_{2j} represents the school-dependent deviation. Its variance is of particular interest, as it expresses the extent to which the effect of schooling differs between schools. When fitting this model, the variances of R_{ij} and U_{0j} are estimated, as is the covariance between U_{0j} and U_{2j}.

In order to establish whether the effect of schooling is dependent on a third variable (e.g. SES), this should be modelled as an interaction effect between this variable (denoted as q) and the effect of schooling. The model then becomes:

$$Y_{ij} = \gamma_{00} + \beta_{10}(x_{ij} - x_0) + \gamma_{20} z_{ij} + \beta_{30}\, q_{ij} + \beta_{40} z_{ij} q_{ij} + U_{0j} + U_{2j} z_{ij} + R_{ij} \quad (4)$$

The interaction may indicate that the effect of schooling is stronger for high-SES students. The coefficient β_3 expresses the main effect of SES (the extent to which high-SES students achieve higher scores), and β_4 expresses the interaction effect of SES and grade level (the extent to which the effect of being in a higher grade is stronger for high SES students). The following sections present the research questions and descriptions of the data, variables and the models fitted in the analysis.

Research questions

The analyses aim to answer the following research questions:

1. What is the effect of one extra year of schooling on student achievement?
2. What proportion of the difference in achievement between two grades is accounted for by schooling?
3. To what extent does the effect of one extra year of schooling vary between schools?
4. To what extent does the effect interact with student and school characteristics?

Data

The data sets analysed in this study derive from the Third International Mathematics and Science Study, usually referred to as TIMSS.[1] These data relate to primary and secondary education in 45 countries, in which mathematics and science tests were administered to more than half a million students. Most countries collected their data in May and June of 1995, but the countries in the southern hemisphere schedule tested in late 1994. The analyses presented here relate to students in primary education, often designated as TIMSS Population 1. Twenty-six countries participated in the primary education component of TIMSS, which focused on the two grades with the largest proportion of nine-year-olds. In most countries, these students were in the third and fourth grades (Gonzalez & Smith, 1997; Martin *et al.*, 1997; Mullis *et al.*, 1997). The analyses focus exclusively on eight countries: Cyprus, Greece, England, Iceland, Japan, Norway, Scotland and Singapore. For these countries, applying the regression-discontinuity approach presents no significant problems, as the percentage of delayed students does not exceed 5%. In two countries, Kuwait and Israel, application of the regression-discontinuity approach is not possible, as tests were administered in only one grade. The main difficulty for the other countries relates to the cut-off point (date of birth) that determines whether a student is placed in a higher or lower grade. The regression-discontinuity approach requires strict adherence to the cut-off point. Inspection of the data revealed that, in most countries, assignment to grade level does not adhere fully to a nation-wide cut-off point. In some cases, this may be (partly) due to regional variations (for example, Australia and the United States), but grade retention is probably the main cause. The school careers of a small number of students are accelerated through placement in a higher grade than is typical for their age. In addition, information on the date of birth is missing for some students.

If the degree of misclassification is not excessive, it is still possible to obtain reliable effect estimates. If the percentage of misclassified participants does not exceed 5%, it is best to exclude them (Judd & Kennedy, 1981; Trochim, 1984; Shadish *et al.*, 2002). The percentage of delayed students in the countries addressed in the analyses is always below 5%. For some countries, however, the percentage of 'correct classifications' is slightly below 95%. See Table 1 for details. The analyses focus exclusively on students with standard school careers.

Variables

The dependent variables in the analysis are student achievement in mathematics and science. The student data files include six types of achievement scores. The ones addressed in this analysis are the 'international proficiency scores'.[2] The user guide for the TIMSS international database (Gonzalez & Smith, 1997) recommends using these scores for both international and within-country comparisons. Table 2 shows the mean and standard deviation for mathematics and science achievement for each country, by grade.

Table 1. Basic statistics per country

		School careers				Sample size	
	Cut-off Point	standard	delayed	accelerated	unknown	students	schools
Cyprus	1 March	92.4%	4.5%	2.0%	1.0%	6,169	187
England	1 Sept.	95.6%	1.1%	0.8%	2.5%	5,960	132
Greece	1 April	94.1%	4.2%	0.9%	0.9%	5,656	175
Iceland	1 Jan.	97.5%	0.8%	0.5%	1.2%	3,373	147
Japan	1 April	98.0%	0.5%	0.6%	0.9%	8,442	142
Norway	1 Jan.	97.3%	1.1%	0.9%	0.8%	4,361	140
Scotland	1 March	94.3%	3.4%	1.7%	0.6%	6,076	154
Singapore	1 Jan.	96.8%	2.6%	0.4%	0.2%	13,598	191

It is important to note that, in each country, the shapes of the frequency distributions of both the mathematics and science achievement scores within each grade are close approximations of the standard normal distribution. There are no signs of any floor or ceiling effects that may bias the effect estimates. The distributions of the science achievement scores of students in lower and upper grades ('Year 4' and 'Year 5') of the English sample, as displayed in Figures 1A and 1B, illustrate this point.

Grade level and date of birth are crucial independent variables in the analyses. Grade level[3] was re-coded to assign scores of zero to students in the lower grade and scores of one to students in the higher grade. The variable that denotes a student's date of birth is based on year and month of birth.[4] Each date was transformed into a single number. For example, a student born in March 1985 received a score of 85.25, and a student born in April received a score of 85.33. The cut-off value was then subtracted from these scores, giving each of the oldest students in the lower

Table 2. Means and standard deviations per grade for mathematics and science

	MATHEMATICS				SCIENCE			
	lower grade		upper grade		lower grade		upper grade	
	mean	*s.d.*	*mean*	*s.d.*	*mean*	*s.d.*	*mean*	*s.d.*
Cyprus	432.8	76.9	504.7	85.7	416.9	72.3	476.7	75.8
England	455.4	87.5	512.6	91.3	497.3	100.6	551.3	96.0
Greece	429.5	85.1	493.1	89.3	448.1	82.1	498.9	82.7
Iceland	410.6	67.2	473.9	72.2	436.1	81.9	505.2	84.7
Japan	538.8	74.9	597.7	80.7	522.7	72.2	574.4	72.1
Norway	421.6	71.5	502.7	74.0	451.0	89.8	531.1	86.2
Scotland	458.4	79.1	520.6	89.1	483.5	94.9	535.5	93.2
Singapore	552.2	100.0	624.6	104.2	488.1	98.9	546.7	97.3

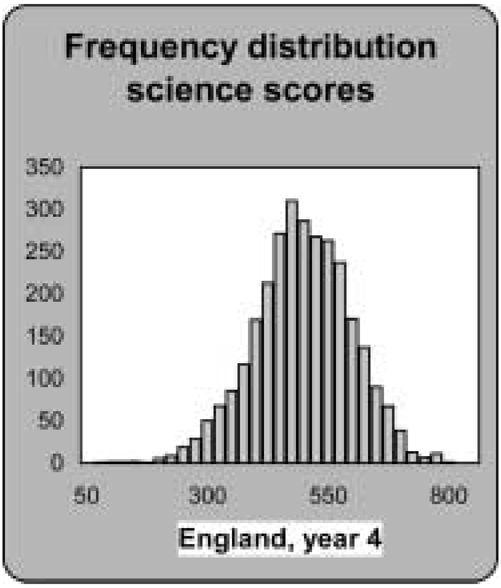

Figure 1A

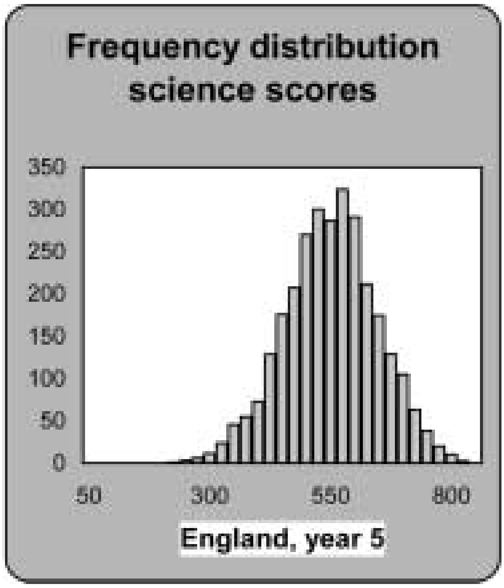

Figure 1B

grade (the comparison group) a score of zero. For the English students in the analysis, the cut-off value was 85.75. Students born in September 1985 or later were placed in the lower grade, and the ones born earlier were placed in the higher grade. Table 3 illustrates the transformation of the original birth dates to the scores used in the analyses, for the English sample.

The analyses also assess the extent to which the effect of schooling differs for certain groups of schools and students. The student characteristics taken into account are sex and the number of books in the student's home.[5] This information derives from a questionnaire administered to the students in addition to the science and mathematics tests. The school variables included in the analyses are the average number of books in the student's home, by school, and the gender composition of the school. The latter variable applies only to the analyses relating to Singapore. This was the only country in which a substantial percentage of the students attended either boys' schools or girls' schools (6.3% and 7.9%, respectively).

Sex was re-coded to assign scores of zero to boys and scores of one to girls. The number of books in a student's home is a five-category variable (0–10; 11–25; 26–100; 101–200; more than 200). Information on this variable was missing for Japan. The variable was re-coded such that a score of zero denotes the middle category (26–100). The two lower categories were coded –2 and –1, while the higher categories were coded +1 and +2. The middle category was the median for four countries, Cyprus, Greece, Scotland and Singapore. For the other three countries (England, Iceland and Norway), the median was one category higher (101–200). As information on the number of books was missing for a substantial number of students (more than 5%) in all countries except Singapore, a dummy variable was created to indicate whether

Table 3. Range of birth dates in the English sample (cut-off = 85.75)

	Year and month	In decimals	Minus cut-off
Upper grade	September 1984	84.75	−1.00
	October 1984	84.83	−.92
	November 1984	84.92	−.83
	December 1984	85.00	−.75
	January 1985	85.08	−.67
	February 1985	85.17	−.58
	March 1985	85.25	−.50
	April 1985	85.33	−.42
	May 1985	85.42	−.33
	June 1985	85.50	−.25
	July 1985	85.58	−.17
	August 1985	85.67	−.08
Lower grade	September 1985	85.75	.00
	October 1985	85.83	.08
	November 1985	85.92	.17
	December 1985	86.00	.25
	January 1986	86.08	.33
	February 1986	86.17	.42
	March 1986	86.25	.50
	April 1986	86.33	.58
	May 1986	85.42	.67
	June 1986	85.50	.75
	July 1986	85.58	.83
	August 1986	85.67	.92

the number of books at home was known. Students for whom the information was missing received scores of one on the dummy variable and scores of zero on the number of books in the home (which denotes the middle category). The other students received scores of zero on the dummy variable.

Models fitted

This analysis fits three models for each country and subject. The first model serves as the reference basis and relates only to the difference in achievement between lower and upper grades, and to the variation of this difference between schools. The second model includes the effect of the student's date of birth, thereby representing the basic regression-discontinuity approach within a multilevel framework, as described in equation (3). The results obtained by fitting this model express the effect of one extra year of schooling, after taking the effect of a student's birth date into account. They also allow for a comparison between the effect of date of birth on achievement and the effect of one extra year of schooling. The third model includes the effects of sex, number of books at home, school average of books at home, gender composition of the school (only for Singapore) and the effects of the dummy variable, which indicates whether information on the number of books at home was missing. Two effects were estimated for each of these variables: the effect on the level of achievement and their interaction with the grade level effect. These interaction effects express whether the effect of one extra year of schooling varies according to sex, number of books at home (individual or school average), gender composition or with the dummy variable that indicates whether information on books at home is missing. The third model also includes an interaction effect of birth date with grade. A significant effect would imply that the effect of birth date on achievement differs between grades. Additional analyses were conducted to explore the existence of curvilinear relationships between age and achievement. The results are presented in the appendix.

Appropriate estimation of effects requires that the TIMSS sample design be taken in account in the data analyses. This can be achieved by using the appropriate sampling weights, which ensure the proper and proportional representation of the various subgroups constituting the sample in the computation of population estimates. The weight applied in the analyses is the one recommended in the user guide for the TIMSS international database (Gonzalez & Smith, 1997) when performing significance tests.[6]

Findings

Tables 4–7 present the results of the analyses. The tables summarise the output obtained with the MLwiN-software (Rasbash *et al.*, 2000). The following section will discuss the findings and their implications in more detail. This section merely draws attention to several general conclusions that can be derived from the tables in a straightforward fashion. The analyses of the curvilinear relationships between age and achievement indicate that a linear function provides the most adequate description of the relationship in the data sets that were analysed. Introducing quadratic and cubic terms fails to provide statistically significant results. See the appendix for more details. The most plausible explanation for this is the restricted age range (two years) of the students in this study. Curvilinear relationships seem most likely in samples over a wider age range.

Model 2 reveals the significant effects of grade level and date of birth on both mathematics and science within each country (see Tables 4 and 5). The effect of date of birth is negative in all cases, implying an advantage for the older students within each grade (a higher birth date denotes a younger age). The grade-level effect is always positive, suggesting that student achievement is significantly higher in the upper grade, after controlling for date of birth. In other words, an extra year of schooling has a positive effect on mathematics and science achievement in each of the eight countries included in the analyses. The analyses further reveal a significant variation of the grade level effect between schools for both subjects in every country, indicating that the effect of education differs from school to school. Furthermore, the analyses produce a consistently negative covariance between the school-level intercept and the grade-level effect. This implies that the effect of an extra year of schooling is relatively low in schools in which the lower grade has a high level of achievement.

The main purpose of Tables 6 and 7 is to show the extent to which the grade level effect interacts with a number of other variables. The significant effects are printed in bold. The consistent lack of statistical significance for the interactions between birth date and grade level is somewhat surprising. While it is logical to expect the effect of age on achievement to become weaker in the upper grades, the results do not support this expectation. Significant *main* effects were found for most of the explanatory variables included in the analyses. These main effects indicate the extent to which the level of achievement in the lower grade coincides with these variables. Neither the finding that, in most countries, girls achieve lower mathematics and science scores, nor that the number of books in the home correlates positively with achievement is surprising. The main effect of the dummy-variable indicating missing information on the number of books at home is significant and negative for both mathematics and science for both in all countries. This result suggests that the achievement scores of the students for whom this information was not available were below those of students in the middle category (26–100 books).

The interaction effects are more interesting, as they indicate the extent to which the effect of one extra year of schooling differs, for example, for boys and girls or for students with few or many books at home. Significant interactions of the grade-level effect with sex were found for mathematics in Iceland and Scotland and for science in Japan. With regard to mathematics, the interactions indicate that, in Iceland and Scotland, the effect of one extra year of schooling is greater for girls than it is for boys. For science achievement, the interaction effect indicates the opposite in Japan. Significant interactions of the grade-level effect with the number of books in the home were found in Iceland, Norway, Scotland and Singapore. They all imply that the effect of one extra year of schooling increases with the number of books at home. The significant interaction effects for Iceland and Scotland relate to mathematics, for Singapore to science and for Norway to both subjects. No significant interaction effects were found for the school mean of books at home. In Singapore, however, the grade-level effect was considerably stronger in girls' schools than it was in other types of schools. The dummy variable indicating whether the information on the number of books at home was missing produced some significant interactions with the grade level effect as well. This result indicates that the grade-level effect for these students differed from the effect for students in the middle category (101–200 books).

Answering the research questions

This section discusses the findings in more detail and links them to the research questions presented earlier.

Table 4. Mathematics, Model 1 and 2 (effects of grade level and date of birth)

		Fixed effects						Random effects at school level						Random effect at student level	
	Model	Intercept		Grade level		Date of birth		Variance of the intercept		Variance of grade level effect		Covariance		Variance of the intercept	
Cyprus	1	435.38	(2.81)	70.98	(3.32)	–	–	902.33	(138.76)	723.04	(240.15)	−195.54	(144.42)	5564.33	(231.70)
	2	448.59	(3.53)	40.39	(5.91)	−30.40	(4.28)	905.49	(136.46)	722.21	(237.01)	−191.12	(142.62)	5495.55	(224.19)
England	1	454.60	(4.10)	57.04	(3.48)	–	–	1901.99	(349.71)	890.10	(211.61)	−424.05	(190.43)	6125.60	(194.55)
	2	470.50	(4.43)	21.98	(4.69)	−35.23	(3.65)	1889.49	(351.20)	871.72	(210.58)	−421.99	(190.22)	6033.59	(193.29)
Greece	1	428.70	(4.12)	67.37	(4.03)	–	–	2559.89	(342.19)	2064.47	(342.84)	−1226.59	(271.98)	5120.20	(202.17)
	2	438.27	(4.38)	44.88	(5.32)	−22.85	(3.51)	2555.74	(314.59)	2089.84	(300.65)	−1236.64	(249.34)	5080.10	(98.57)
Iceland	1	410.20	(2.65)	62.27	(2.91)	–	–	444.13	(108.00)	279.10	(161.94)	−145.18	(94.91)	4501.19	(263.55)
	2	420.95	(3.18)	39.15	(5.07)	−23.22	(4.38)	439.29	(106.13)	279.59	(166.17)	−148.57	(94.64)	4462.11	(262.09)
Japan	1	538.22	(1.87)	58.70	(1.88)	–	–	285.89	(58.18)	102.98	(58.02)	−27.30	(45.11)	5760.70	(90.14)
	2	549.99	(2.30)	32.53	(3.77)	−26.47	(3.34)	291.26	(73.62)	109.74	(51.94)	−32.96	(45.09)	5703.42	(148.47)
Norway	1	423.00	(2.97)	80.99	(3.33)	–	–	879.12	(158.43)	879.30	(185.62)	−393.73	(134.93)	4452.17	(178.81)
	2	433.35	(3.44)	58.10	(5.05)	−22.47	(3.71)	884.79	(146.01)	882.95	(183.20)	−399.92	(130.69)	4414.04	(97.53)
Scotland	1	457.31	(3.52)	64.97	(3.27)	–	–	1599.64	(300.22)	1022.39	(184.85)	−110.86	(157.79)	5121.38	(135.76)
	2	470.28	(3.85)	35.71	(4.71)	−28.90	(3.42)	1594.48	(216.72)	933.77	(173.27)	−92.32	(140.38)	5068.12	(94.27)
Singapore	1	543.34	(4.60)	73.88	(5.11)	–	–	3821.33	(433.01)	4560.73	(597.81)	−2053.80	(426.86)	6384.93	(217.93)
	2	552.06	(4.72)	55.74	(5.63)	−18.27	(2.41)	3765.04	(407.86)	4503.75	(504.48)	−2008.35	(358.70)	6359.71	(78.24)

NB: The figures in brackets denote the standard errors

Table 5. Science, Model 1 and 2 (effects of grade level and date of birth)

		Fixed effects						Random effects at school level						Random effect at student level	
	Model	Intercept		Grade level		Date of birth		Variance of the intercept		Variance of grade level effect		Covariance		Variance of the intercept	
Cyprus	1	420.12	(2.43)	57.54	(2.92)	–	–	661.19	(114.39)	520.22	(129.92)	–43.59	(77.05)	4589.93	(212.65)
	2	432.49	(3.06)	28.96	(5.48)	–28.37	(4.13)	672.22	(114.13)	521.89	(131.24)	–49.85	(77.23)	4529.74	(206.46)
England	1	496.08	(4.40)	53.64	(3.74)	–	–	2153.08	(293.08)	987.45	(260.30)	–650.48	(246.93)	7672.03	(256.15)
	2	512.05	(4.97)	18.44	(5.46)	–35.37	(4.38)	2130.38	(290.87)	947.99	(250.04)	–639.48	(241.92)	7582.66	(255.63)
Greece	1	446.57	(3.85)	54.42	(3.61)	–	–	2218.45	(319.22)	1593.58	(251.41)	–926.50	(237.19)	4641.19	(210.59)
	2	456.52	(4.07)	31.03	(5.23)	–23.77	(3.68)	2215.43	(313.57)	1598.24	(253.25)	–927.17	(235.66)	4598.62	(209.39)
Iceland	1	435.37	(3.32)	68.87	(3.51)	–	–	745.61	(157.72)	428.99	(205.52)	–319.41	(161.49)	6388.19	(322.24)
	2	448.12	(3.92)	41.45	(6.01)	–27.53	(4.92)	734.63	(156.48)	441.51	(208.57)	–326.28	(161.61)	6332.13	(322.15)
Japan	1	522.54	(1.68)	51.57	(1.99)	–	–	219.43	(46.93)	213.62	(62.64)	–85.88	(41.54)	4970.97	(123.14)
	2	534.60	(2.00)	24.73	(3.36)	–27.15	(2.84)	219.59	(48.25)	207.34	(62.35)	–81.12	(41.41)	4912.19	(123.38)
Norway	1	450.14	(3.45)	81.12	(4.22)	–	–	1123.03	(218.31)	1467.83	(310.35)	–823.72	(229.03)	6734.81	(269.26)
	2	465.51	(4.05)	47.11	(6.27)	–33.41	(4.55)	1132.78	(194.83)	1460.41	(289.75)	–827.29	(201.24)	6650.51	(146.96)
Scotland	1	482.53	(4.43)	54.68	(3.39)	–	–	2621.40	(470.56)	975.59	(177.08)	–679.82	(247.99)	6519.86	(188.15)
	2	497.05	(5.03)	21.92	(5.50)	–32.38	(4.34)	2621.37	(469.08)	897.64	(165.58)	–670.55	(244.54)	6452.75	(185.10)
Singapore	1	479.41	(4.59)	60.17	(4.61)	–	–	3834.34	(400.50)	3654.51	(478.31)	–1886.93	(371.73)	5805.67	(199.39)
	2	486.17	(4.62)	46.10	(5.29)	–14.16	(2.40)	3790.33	(394.19)	3615.10	(473.01)	–1850.95	(366.92)	5790.88	(199.29)

NB: The figures in brackets denote the standard errors

Table 6. Mathematics, Model 3 (interactions with grade level)

FIXED EFFECTS	CYPRUS		ENGLAND		GREECE		ICELAND	
Intercept	**458.11**	(4.13)	**453.05**	(6.69)	**451.59**	(4.71)	**414.01**	(6.85)
Grade level	**41.63**	(6.21)	**22.38**	(6.98)	**51.09**	(5.29)	**23.72**	(8.32)
Date of birth	**–25.49**	(5.39)	**–39.16**	(5.17)	**–16.74**	(4.95)	**–20.63**	(5.79)
Date of birth*	–3.11	(7.08)	9.99	(8.14)	–.40	(7.74)	–4.45	(8.07)
Girls	**–8.76**	(2.98)	**–7.76**	(2.76)	–3.58	(2.87)	**–17.29**	(3.24)
Girls*	.14	(3.99)	–.04	(4.87)	–.72	(4.28)	**16.09**	(4.55)
Number of books at home	**6.37**	(1.37)	**13.64**	(1.21)	**11.78**	(1.28)	**5.98**	(1.55)
Number of books at home*	.85	(1.87)	2.01	(1.94)	3.20	(2.09)	**5.26**	(2.17)
School mean books at home	**17.84**	(8.73)	**39.97**	(8.23)	2.01	(11.85)	**21.25**	(8.65)
School mean books at home*	.56	(10.34)	–3.45	(7.89)	17.93	(9.59)	4.00	(10.00)
Books at home unknown	**–46.34**	(3.35)	**–51.44**	(7.95)	**–49.45**	(3.52)	**–47.57**	(6.77)
Books at home unknown*	**–17.63**	(5.85)	9.69	(10.26)	**–21.91**	(6.03)	2.86	(10.33)
Boys' school	–	–	–	–	–	–	–	–
Boys' school*	–	–	–	–	–	–	–	–
Girls' school	–	–	–	–	–	–	–	–
Girls' school*	–	–	–	–	–	–	–	–
RANDOM EFFECTS, SCHOOL LEVEL								
Intercept	**815.20**	(125.27)	**1469.02**	(388.18)	**2152.13**	(314.12)	**342.07**	(88.63)
Grade level	**614.52**	(191.38)	**822.57**	(174.75)	**1878.74**	(308.29)	**233.35**	(107.23)
Covariance	**–239.36**	(133.18)	**–386.22**	(188.52)	**–1276.25**	(254.35)	**–150.97**	(80.72)
RANDOM EFFECTS, STUDENT LEVEL								
Intercept	**5137.48**	(205.10)	**5632.22**	(178.42)	**4495.33**	(184.05)	**4230.26**	(106.38)

Table 6. (*Continued*)

FIXED EFFECTS	CYPRUS		NORWAY		SCOTLAND		SINGAPORE	
Intercept	**553.44**	(2.87)	**425.16**	(7.33)	**458.71**	(5.07)	**547.43**	(4.14)
Grade level	**35.29**	(3.86)	**32.32**	(9.66)	**28.27**	(5.56)	**53.90**	(6.52)
Date of birth	**–29.86**	(4.14)	**–21.95**	(5.38)	**–24.53**	(5.04)	**–23.15**	(3.44)
Date of birth*	7.50	(5.84)	2.36	(7.30)	–3.24	(8.00)	7.98	(4.79)
Girls	–3.92	(2.36)	**–16.55**	(3.23)	**–7.10**	(2.73)	–.55	(2.04)
Girls*	–3.95	(3.30)	8.40	(4.62)	**6.90**	(3.53)	–.54	(3.02)
Number of books at home	–	–	**9.27**	(1.39)	**12.59**	(1.22)	**13.30**	(1.01)
Number of books at home*	–	–	**5.74**	(2.02)	**3.46**	(1.70)	2.38	(1.35)
School mean books at home	–	–	17.48	(9.64)	**31.00**	(5.86)	**80.68**	(7.80)
School mean books at home*	–	–	21.37	(11.59)	8.99	(6.02)	–11.70	(10.81)
Books at home unknown	–	–	**–52.40**	(5.87)	**–49.60**	(6.69)	**–75.38**	(12.85)
Books at home unknown*	–	–	14.29	(10.67)	**–26.20**	(10.70)	–18.05	(23.90)
Boys' school	–	–	–	–	–	–	–8.38	(8.90)
Boys' school*	–	–	–	–	–	–	11.32	(14.96)
Girls' school	–	–	–	–	–	–	–14.39	(9.97)
Girls' school*	–	–	–	–	–	–	**42.18**	(13.70)
RANDOM EFFECTS, SCHOOL LEVEL								
Intercept	**293.33**	(58.82)	**629.15**	(127.68)	**1044.49**	(257.45)	**1779.99**	(220.05)
Grade level	**110.00**	(58.40)	**652.72**	(156.37)	**845.53**	(161.28)	**3891.35**	(504.79)
Covariance	**–34.30**	(45.62)	**–344.98**	(116.72)	**–270.48**	(123.56)	**–1734.23**	(293.45)
RANDOM EFFECTS, STUDENT LEVEL								
Intercept	**5694.18**	(89.10)	**4070.73**	(164.49)	**4604.53**	(112.29)	**6051.71**	(199.84)

The effects marked with an asterisk (*) indicate the interaction with grade level.
The figures in brackets denote the standard errors.
The effects printed in bold are significant at $\alpha < .05$ (two-tailed).

Table 7. Science, Model 3 (interactions with grade level)

FIXED EFFECTS	CYPRUS		ENGLAND		GREECE		ICELAND	
Intercept	**442.10**	(3.41)	**487.63**	(7.76)	**476.65**	(4.10)	**430.49**	(8.40)
Grade level	**30.67**	(5.56)	**15.87**	(8.24)	**34.72**	(4.91)	**40.54**	(10.05)
Date of birth	**–23.57**	(5.76)	**–37.30**	(5.72)	**–21.24**	(4.67)	**–21.97**	(6.90)
Date of birth*	–2.90	(6.68)	5.87	(8.41)	6.58	(6.44)	–10.06	(9.61)
Girls	**–8.63**	(2.35)	–5.33	(3.67)	**–11.23**	(2.90)	**–10.86**	(3.86)
Girls*	–2.70	(3.40)	–1.72	(4.97)	.32	(4.16)	–6.71	(5.42)
Number of books at home	**8.95**	(1.22)	**20.69**	(1.41)	**10.80**	(1.29)	**9.88**	(1.85)
Number of books at home*	–.91	(1.70)	–.11	(2.09)	3.39	(2.20)	4.10	(2.59)
School mean books at home	**17.55**	(7.21)	**44.58**	(9.18)	7.97	(12.20)	**27.05**	(10.70)
School mean books at home*	4.18	(9.02)	2.31	(8.56)	9.18	(9.12)	.10	(12.14)
Books at home unknown	**–48.58**	(3.44)	**–65.83**	(9.86)	**–53.21**	(3.40)	**–48.56**	(8.07)
Books at home unknown*	–8.39	(5.39)	**33.25**	(13.29)	**–14.15**	(7.00)	.24	(12.31)
Boys' school	–	–	–	–	–	–	–	–
Boys' school*	–	–	–	–	–	–	–	–
Girls' school	–	–	–	–	–	–	–	–
Girls' school*	–	–	–	–	–	–	–	–
RANDOM EFFECTS, SCHOOL LEVEL								
Intercept	**547.59**	(96.32)	**1526.52**	(359.67)	**1696.67**	(284.52)	**564.89**	(138.07)
Grade level	**473.86**	(116.83)	**1004.30**	(243.94)	**1475.87**	(229.07)	**379.36**	(160.12)
Covariance	**–114.72**	(67.78)	**–750.92**	(243.46)	**–914.53**	(220.36)	**–300.20**	(125.95)
RANDOM EFFECTS, STUDENT LEVEL								
Intercept	**4153.32**	(186.48)	**6842.85**	(223.55)	**4020.53**	(212.28)	**6003.30**	(150.98)

Table 7. (*Continued*)

FIXED EFFECTS	JAPAN		NORWAY		SCOTLAND		SINGAPORE	
Intercept	**538.35**	(2.56)	**450.35**	(9.45)	**474.47**	(5.75)	**483.79**	(3.93)
Grade level	**31.02**	(3.63)	**31.88**	(14.06)	**22.50**	(5.89)	**47.63**	(5.83)
Date of birth	**–31.96**	(3.90)	**–34.38**	(7.03)	**–24.55**	(6.01)	**–18.14**	(3.38)
Date of birth*	10.81	(5.73)	5.50	(9.37)	–9.45	(8.21)	6.22	(4.91)
Girls	–3.24	(2.33)	**–10.77**	(3.84)	–1.92	(2.79)	**–10.06**	(2.08)
Girls*	**–10.23**	(3.31)	–.49	(5.57)	–3.28	(4.02)	–5.19	(3.29)
Number of books at home	–	–	**9.25**	(1.94)	**18.69**	(1.24)	**14.32**	(.89)
Number of books at home*	–	–	**8.80**	(2.61)	–1.27	(1.77)	**3.75**	(1.26)
School mean books at home	–	–	**26.07**	(12.43)	**44.28**	(7.13)	**82.96**	(6.87)
School mean books at home*	–	–	8.74	(16.49)	2.34	(6.24)	–15.71	(8.98)
Books at home unknown	–	–	**–70.58**	(7.89)	**–58.90**	(7.94)	**–75.28**	(10.15)
Books at home unknown*	–	–	13.11	(12.47)	–7.59	(10.64)	–6.59	(18.67)
Boys' school	–	–	–	–	–	–	5.12	(7.64)
Boys' school*	–	–	–	–	–	–	1.02	(15.57)
Girls' school	–	–	–	–	–	–	–5.34	(9.87)
Girls' school*	–	–	–	–	–	–	**38.25**	(11.58)
RANDOM EFFECTS, SCHOOL LEVEL								
Intercept	**220.56**	(48.59)	**902.78**	(189.27)	**1349.93**	(292.08)	**1534.71**	(206.43)
Grade level	**208.75**	(61.75)	**1280.63**	(288.01)	**852.97**	(162.47)	**3061.40**	(387.93)
Covariance	**–83.84**	(41.48)	**–773.54**	(220.76)	**–641.12**	(157.97)	**–1487.49**	(248.03)
RANDOM EFFECTS, STUDENT LEVEL								
Intercept	**4887.76**	(123.03)	**6146.80**	(259.73)	**5821.44**	(159.13)	**5387.38**	(177.93)

The effects marked with an asterisk (*) indicate the interaction with grade level.
The figures between brackets denote the standard errors.
The fixed effects printed in bold are significant at $\alpha < .05$ (two-tailed).

The effect of one extra year of schooling on achievement

The grade-level effect for mathematics, when taking the effect of date of birth into account, ranges from 22 points in England to 58 points in Norway (see Table 4, Model 2). Figures 2 and 3 present graphical displays of these outcomes.

The two figures show similar patterns. In both England and Norway, the older students within each grade achieved higher scores, while there was also clear discontinuity at the cut-off point, which denotes the effect of one extra year of schooling. The youngest students in the upper grade achieved considerably higher scores than did the oldest students in the lower grades. This pattern was found for both mathematics and science in every country. Nonetheless, the figures reveal marked differences as well. The effect of being in the upper grade in Norway was more than two and one half times larger than the effect in England (58.10 and 21.98, respectively). In England, the effect of age was about one and one half times as strong as that in Norway (35.23 and 22.47, respectively). The effect of age is the same as the effect of date of birth, but with the sign reversed. The effect of age on mathematics achievement was strongest in England (35.23), while Singapore had the smallest effect (18.27).

The effect of being in the upper grade was smaller for science than it was for mathematics in all countries except Iceland. The effect of age was similar for both subjects. In all countries, the size of both effects on science achievement corresponds closely to the size of the effects on mathematics achievement. Countries with large grade-level effects on mathematics achievement revealed relatively large effects for science as well. The same pattern applies for the effect of age. For science, the largest and smallest effects of being in the upper grade were again found in England (18.44)

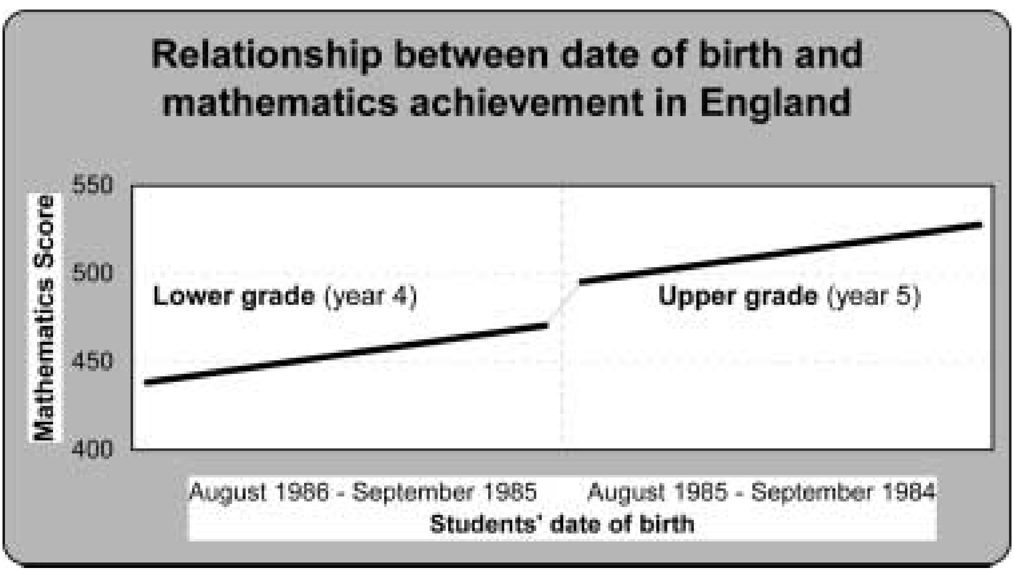

Figure 2. Relationship between date of birth and mathematics achievement in England

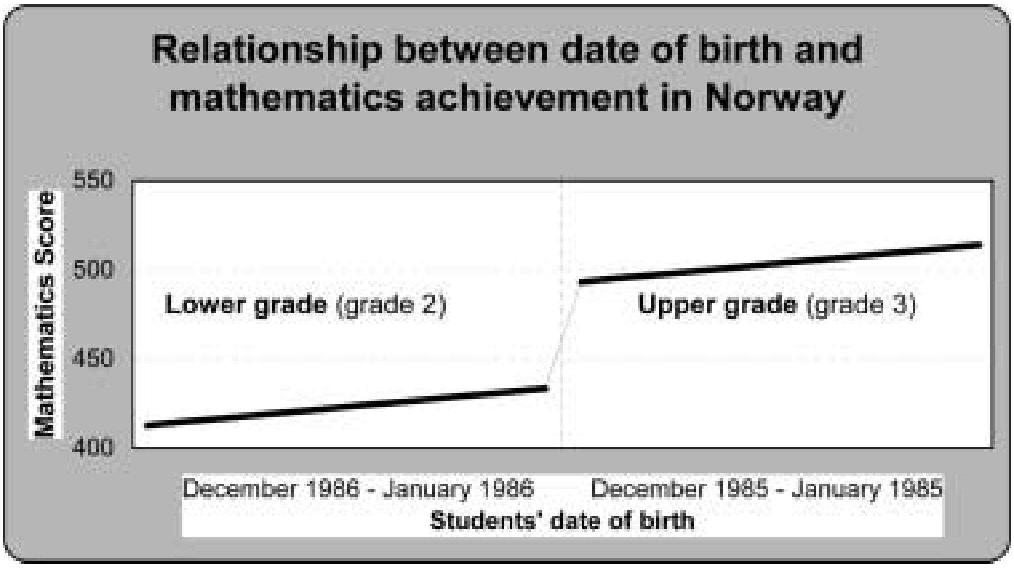

Figure 3. Relationship between date of birth and mathematics achievement in Norway

and Norway (47.11). The effect of age ranged from 14.16 in Singapore to 35.37 in England.

Proportion of difference in achievement between two grades accounted for by schooling

The difference in achievement between students in the upper and lower grades can be seen as consisting of two components, the effect of age and the effect of one extra year of schooling. With regard to mathematics, the latter effect outweighed the effect of age in all countries except England. This implies that, for mathematics, the effect of schooling accounts for more than half of the difference in achievement between both grades in seven countries. Table 8 shows the exact percentages, and Figure 4 presents a graphical display of the sizes of the age effect and the grade-level effect on mathematics achievement.

The relative size of the grade-level effect was somewhat smaller with respect to science (see Figure 5 and Table 8). In three countries (England, Scotland and Japan), this effect accounted for less than half of the difference in achievement between the students in the upper and the lower grade. For both mathematics and science, England presented the lowest percentages (38% and 34%, respectively), and Singapore had the highest (75% and 77%).

Inspection of Figures 4 and 5 and Tables 4 and 5 suggests that the effect of age tends to increase as the grade-level effect decreases. Large age effects, however, can compensate only partially for small grade-level effects. Differences in achievement between students in the upper and the lower grade were explained primarily by the effect of an extra year of schooling.

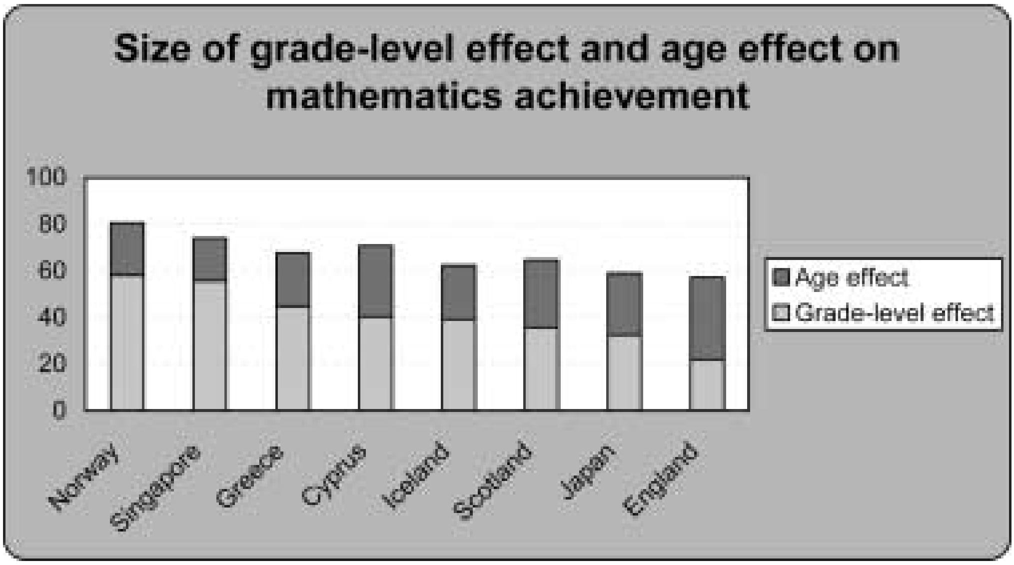

Figure 4. Size of grade-level effect and age effect on mathematics achievement

Variation between schools in the effect of one extra year of schooling

The analyses reveal significant variation of the grade-level effect for both subjects in all eight countries. The variances of the grade-level effects reported in Tables 4 and 5 were transformed into standard deviations, in order to give a better idea of their

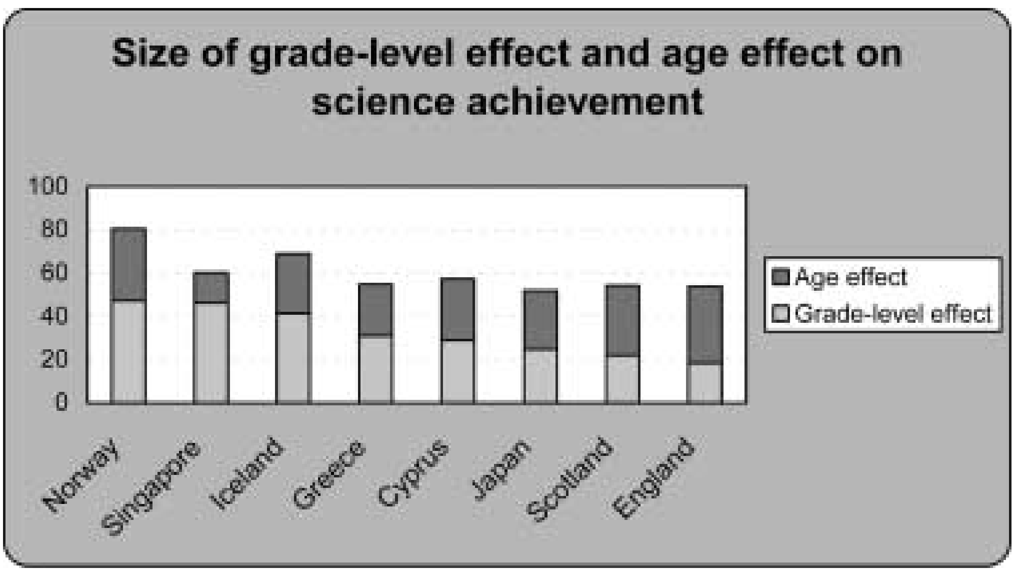

Figure 5. Size of grade-level effect and age effect on science achievement

Table 8. Implications of the findings reported in Tables 4 and 5

	Effect of one year extra schooling as a percentage of the difference between grades		Range of the grade level effect for the middle half of schools		Percentage of schools without a positive effect of grade level		Correlation between grade level effect and level of achievement in lower grade	
	Maths (%)	*Science (%)*	*Maths*	*Science*	*Maths (%)*	*Science (%)*	*Maths*	*Science*
Cyprus	57	51	22–59	14–44	7	10	–.24	–.08
England	38	34	2–42	–2–39	23	27	–.33	–.45
Greece	66	57	14–76	4–58	16	22	–.54	–.49
Iceland	63	60	28–50	27–56	1	2	–.42	–.57
Japan	55	48	25–40	15–34	0	4	–.18	–.38
Norway	72	59	38–78	21–73	2	11	–.45	–.64
Scotland	55	40	15–56	2–42	12	23	–.08	–.44
Singapore	75	77	11–101	6–87	20	22	–.49	–.50

practical implications. The standard deviations were used to estimate the range of the grade/level effects for the middle half of the schools in each country. In a standard normal distribution, half of the scores fall within the range of 0.67 times the standard deviation, both above and below the mean.[7] Table 8 shows the estimated ranges, and Figures 6 and 7 present a graphical display. For both subjects, the range of the grade-level effect was smallest in Japan. The range was also limited in Iceland and Cyprus. Singapore and Greece had the widest range of variation for both mathematics and science. England, Norway and Scotland took up the middle positions.

As shown in Figure 6, the lower bound of the range for England was close to zero. The bars represent the middle half of the schools. In the bottom quarter of the English schools, the effect was therefore barely above zero. This result implies that, in nearly 25% of the schools, the effect of one extra year of schooling on mathematics was negative, when taking the effect of age into account. With regard to science, the lower bound was close to zero in Scotland and was slightly below zero in England (see Figure 7 and Table 8). In Greece and Singapore, the lower bound was also close to zero with regard to science achievement. Table 8 presents estimates of the percentages of schools in each country that had no positive grade-level effects,[8] and Figure 8 provides a graphical display.

In each country, the percentage of schools that had no positive grade-level effects was smaller for mathematics than it was for science. For mathematics, the percentage of schools without positive grade-level effects was very low in Japan, Iceland and Norway. The percentages in Japan and Iceland were also very low with regard to science achievement. In England, Singapore, Greece and Scotland, more

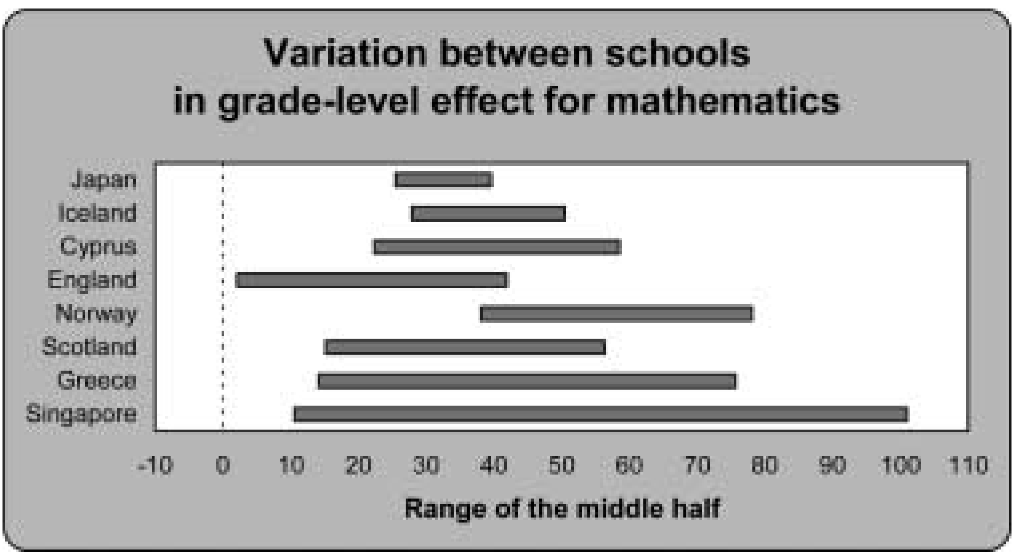

Figure 6. Variation between schools in grade-level effect for mathematics

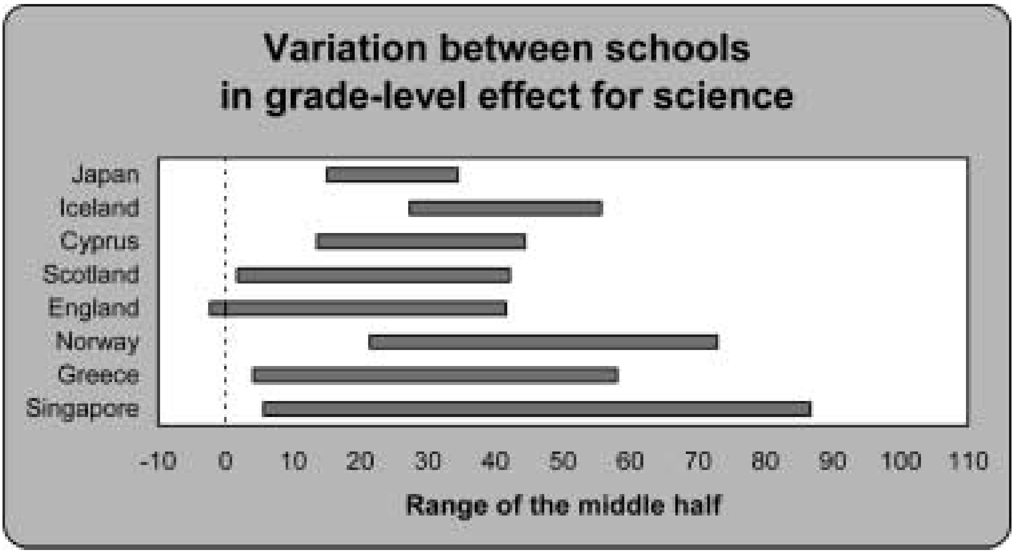

Figure 7. Variation between schools in grade-level effect for science

than 20% of the schools showed no positive grade-level effects for science. In one out of five schools in these countries, an extra year of schooling had no positive effect on science achievement, when controlling for the effect of age. In Cyprus and Norway, the percentages are more moderate, but nevertheless substantial (10% and 11%, respectively). With regard to mathematics, Singapore and England had the highest percentages of schools without a positive grade-level effect (20% and 23%, respectively), while the percentages in Cyprus (7%), Scotland (12%) and Greece (16%) were more limited, but still substantial. In one out of 14 schools in Cyprus, there was no advantage of being in the upper grade with regard to mathematics achievement. In Scotland, this held for one out of eight schools, and in Greece, for one out of six.

The analyses also consistently show a negative covariance between the grade-level effect and the school-level intercept, which implies that the effect of being in the upper grade was stronger in schools that had modest levels of achievement in the lower grade. For each country, the correlations between the size of the grade level effect and the school level intercept were computed according to the variances and covariances obtained when fitting Model 2 (see Tables 4 and 5).[9] Table 8 reports the correlations. For both mathematics and science, half of the correlations were within the range of -.30 and -.50, which may be considered medium to large (Cohen, 1988). Correlations closer to zero were found in Japan and Scotland for mathematics achievement and in Cyprus for both subjects. Correlations of -.50 or stronger were found for mathematics in Greece and for science in Iceland, Norway and Singapore.

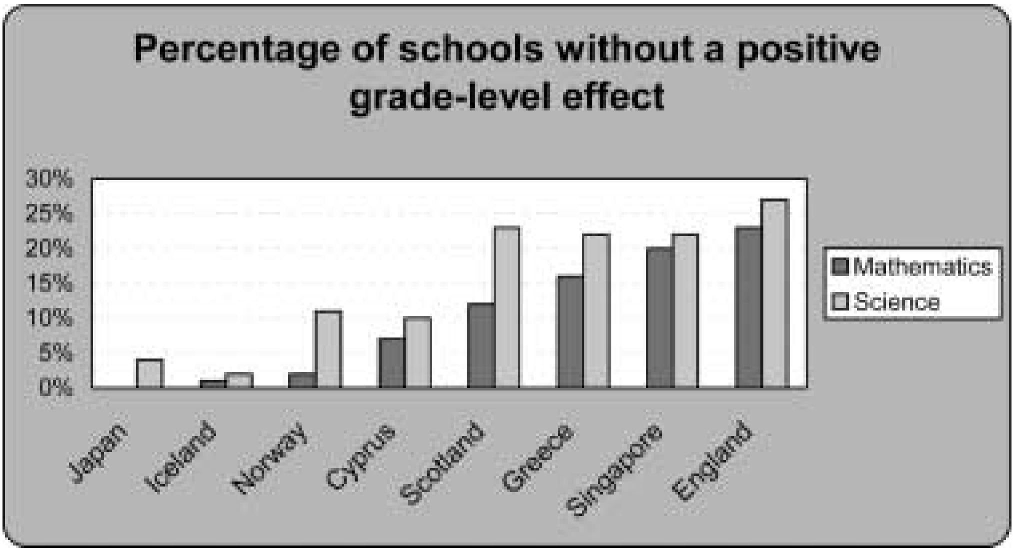

Figure 8. Percentage of schools without a positive grade-level effect

Interaction of the grade-level effect with student and school characteristics

As already mentioned, the analyses reveal a number of significant interaction effects of grade level with other student and school characteristics. These interactions indicate the extent to which the effect of one extra year of schooling is different for certain groups of students or schools. Table 9 displays these differences between boys and girls, and between students with few and many books at home (11–25 and 101–200, respectively).

In Iceland and Scotland, the effect of being in the upper grade on mathematics achievement was stronger for girls than it was for boys. When taking the impact of age into account, the effect of one extra year of schooling was 23.7 for the male students in Iceland, while it was 39.8 for the female students. In Scotland, the effect was 28.3 for boys and 35.2 for girls. A significant interaction of grade level with sex was found in Japan as well, but it involved science achievement and indicated a weaker effect for girls. In Japan, the grade-level effect was 31.0 for boys and 20.1 for girls.

All five significant interaction effects between grade level and the number of books at home imply stronger effects of schooling for students who have many books at home. Three significant interactions related to mathematics, and they were found in Iceland, Norway and Scotland. Significant interactions regarding science achievement were found in Norway and Singapore. The strongest interaction effect between grade level and books at home related to science achievement in Norway. For students who had between 101 and 200 books at home, the grade-level effect was 40.7, whereas it was only 23.1 for Norwegian students who had between 25 and 100 books at home.

Table 9. Grade-level effects by gender and number of books at home (based on the interaction effects reported in Tables 6 and 7)

	MATHEMATICS				SCIENCE			
	Boys	Girls	Books at home 11-25	Books at home 101-200	Boys	Girls	Books at home 11-25	Books at home 101-200
Cyprus	41.6	41.8	40.8	42.5	30.7	28.0	31.6	29.8
England	22.4	22.3	20.4	24.4	15.9	14.2	16.0	15.8
Greece	51.1	50.4	47.9	54.3	34.7	35.0	31.3	38.1
Iceland	**23.7**	**39.8**	**18.5**	**29.0**	40.5	33.8	36.4	44.6
Japan	35.3	31.3	–	–	**31.0**	**20.8**	–	–
Norway	32.3	40.7	**26.6**	**38.1**	31.9	31.4	**23.1**	**40.7**
Scotland	**28.3**	**35.2**	**24.8**	**31.7**	22.5	19.2	20.2	24.8
Singapore	53.9	53.4	51.5	56.3	47.6	42.4	**43.9**	**51.4**

The differences printed in bold are significant at $\alpha < .05$ (two-tailed)

The analyses also reveal that the grade-level effect in Singapore was considerably stronger in girls' schools than it was in other types of schools for both mathematics and science (see Tables 6 and 7). The grade-level effect for mathematics was 96.1 in girls' schools, while it was 53.9 in other schools. For science, the effect of grade level was 85.9 in girls' schools and 47.6 in other schools. The average number of books at home by school produced no significant interactions with the grade-level effect.

Conclusion and discussion

The analyses presented in this article illustrate the usefulness of the regression-discontinuity approach within a multilevel framework for school effectiveness research. In contrast to the more typical forms of data analysis, this approach yields an estimate of the absolute effect of schooling. The regression-discontinuity approach allows strong causal inferences, as it controls for the effect of the criterion that determines assignment to treatment or control groups. Any remaining difference between the groups must therefore be due either to the effect of schooling or to random fluctuation, provided that the relation between the assignment criterion and the independent variable (i.e. age and achievement) is modelled correctly. This approach may also be very convenient, as controlling for prior achievement is—in principal—not required. Such control is essential in standard multilevel analysis; without it, intake differences between schools are bound to cause bias. The results of the regression-discontinuity approach are subject to bias only if the cut-off point coincides with other significant factors. Although this is unlikely in most cases, the effects of schooling as estimated in this study may be biased to some extent. Effects of pre-school provision or peer group effects of being the oldest or youngest in a class may coincide with the cut-off point. The extent to which these and other possibly confounding factors lead to bias in the regression-discontinuities are questions that should be addressed in future research. Testing the validity of the assumption that the average achievement levels are stable across cohorts is another issue for subsequent study.

As this article has shown, assessing the impact of background variables may also be worthwhile from other perspectives. The extent to which the effect of schooling differs between certain groups of schools or students is an especially important issue. For example, the analyses show that, in a number of countries, the effect of one year of schooling was stronger for students with advantaged backgrounds; in Singapore, the effect was considerably stronger in girls' schools than it was in other types of schools. By combining the regression-discontinuity approach with multilevel analysis, this approach facilitates the assessment of differences between schools, producing interesting findings of considerable practical importance. For several countries, one extra year of schooling did not always yield a positive effect on achievement. This disturbing finding calls for an explanation. Floor or ceiling effects in test-score distributions cannot account for this result. In each country, the distributions within each grade were close approximations of the standard normal distribution (see Figures 1A and 1B).

At this point, it is only possible to put forward a few tentative suggestions. Schools may make a deliberate decision to focus on subjects other than mathematics and science in the upper grades. It is conceivable that, in some schools, this will unintentionally lead to higher achievement levels in the lower grades as compared to the upper grades. It may also be the case that, in some schools, teachers in the lower grades do an excellent job, while the teaching in the upper grades is quite poor. The regression-discontinuity approach can also not rule out the possibility that adaptive teaching may produce small grade-level effects. Careful consideration of a student's development before presenting new subject matter may cause some students in the upper grades to perform at a lower level than others in the lower grades.

In the present study, it was only possible to investigate the difference between two consecutive grade levels. This is the most basic regression-discontinuity model. Were appropriate data available, it would be possible to fit models with multiple cut-off points. This would require test scores that cover a wider range of consecutive grades, but that could still be reduced to a common scale. For example, equating the tests according to Item Response Theory would make it possible to estimate the effects of several years of schooling, whereas the present study could assess only the effect of one year. One of the most striking results is the percentages of school having negative grade-level effects. In a remarkable number of schools, students in the lower grades outperformed those in the upper grades, after controlling for the effect of age. The presence of so many negative schooling effects would not be likely in a study covering a wider range of grades.

Such a study would also allow a more extensive analysis of the effects of age on achievement. Somewhat surprisingly, this study revealed no significant interactions of age with grade level, although it seems plausible that the effect of age would decline in higher grade levels. The lack of significant interactions between age and grade level may be due to the limited range of grades addressed in the data. A strong correlation between age (or date of birth) and achievement is not necessary for establishing the effect of schooling using the regression-discontinuity approach. If the correlation between age and achievement within grades is zero, the regression-discontinuity approach reduces to a comparison between mean achievement in the lower and the upper grades. This approximates random assignment, as there is no relation between the assignment criterion and the outcome measure (Trochim, 1984, pp. 82–83).

The regression-discontinuity approach may also be useful for detecting the effect of schooling on such non-cognitive measures as attitudes, behaviour and social competencies. Schools are often expected to contribute to the development of citizenship and social competencies. Whether schools actually accomplish this, however, is still an open question. Ten Dam and colleagues (2003) report that meta-analyses regarding the effects of projects aimed at developing these skills and attitudes are far from encouraging, and that differences between schools have been found to be very small with regard to citizenship and social competencies (Van der Wal, 2004). Nonetheless, school processes that are not part of the official curriculum are often believed to have a strong impact on the development of students' attitudes and social competencies (Giroux, 1983). The fact that differences between schools are small, or that relatively short-term projects have little impact, does not preclude a considerable effect of education in general. With regard to cognitive achievement measures, this study clearly presents a different perspective on school effects as compared to the results from standard multilevel analyses. The findings indicate that, in the eight countries included in the analyses, schooling accounts for an average of 60% of the increase in mathematics achievement between two grade levels. For science, schooling accounts for 53% of the difference. This is somewhat less than the proportion of two thirds reported by Cahan and Davis (1987), whose findings are nonetheless within the range of the per-country percentages reported in this article (see Table 8).

There is probably just one feasible alternative method for assessing the absolute effect of schooling: comparing the development of students during the school year with their development during the summer holidays. Findings from studies that have applied this method also report substantial effects of schooling (Heyns, 1978; Cooper *et al.*, 1996; Lindahl, 2001). Furthermore, they confirm that the effect is stronger for students from advantaged backgrounds. The practical advantage of the regression-discontinuity approach is that it does not require longitudinal data to assess the effect of schooling. The most important issue concerning the effect of schooling is ultimately how it manifests itself; as studying this issue requires long-term follow-up data, the practical advantages of cross-sectional data are evident. Collection of longitudinal data requires more time, effort and money than the collection of cross-sectional data. Moreover, longitudinal data frequently suffer from bias due to selective attrition. If these problems can be addressed, however, longitudinal data provide the best possible information for assessing the effects of schooling. Nonetheless, one of the aims this article has been to show that cross-sectional data may allow much stronger conclusions about the effects of schooling than is usually presumed.

The 'misclassification' of students in another important problem with the regression-discontinuity approach. In all countries, other factors besides date of birth influence assignment to a particular grade. In only one third of the countries that participated in TIMSS is the proportion of 'misclassified' students small enough to obtain reliable estimates of the effect of schooling, without considering other factors. Future research should identify other grade assignment factors in addition to date of birth; the greatest challenge will be to develop valid methods for considering their effects when assessing the effect of schooling. Grade retention is undoubtedly one of the main reasons that assignment to grade level does not always agree with a cut-off point based on date of birth. Regional variations within countries may also be an important factor with regard to the cut-off point.

Notes

1. The database can be downloaded from the internet. Go to http://timss.bc.edu
2. The variable names are AIMATSCR (International Mathematics Achievement Score—Population 1) and AISCISCR (International Science Achievement Score—Population 1).

3. The variable name is IDGRADER (Grade).
4. The variable names are ASBGBIRM (month of birth) and ASBGBIRY (year of birth).
5. The variable names are ASBGSEX (student's sex) and ASBGBOOK (number of books in student's home).
6. The variable name is HOUWGT (House Weight). Its sum adds up to the sample size within each country.
7. For example, the variance between schools of the grade-level effect on mathematics for England in Model 2 was 871.72 (see Table 4), which corresponds with a standard deviation of 29.52. The fixed grade-level effect (the mean effect across schools) was 21.98. The lower bound of the range for the middle half of the schools then became: 21.98 – .67*29.52 = 2.20 and the upper bound became 41.76.
8. To illustrate the computation of these percentages, consider the example of England. The standard deviation of the grade-level effect on mathematics is 29.52. The fixed effect represents the mean effect across schools, and is equal to 21.98. This means that an effect of zero is .745 times the standard deviation below the mean effect. In a standard normal distribution, this implies that the effect is below zero for 23% of the English schools.
9. In England, the covariance for mathematics in Model 2 was –421.99 (see Table 4) and the school-level variances were 1889.49 and 871.72. The correlation thus became: –421.99/ $\sqrt{(1889.49*871.72)}$ = –.33.

Notes on contributor

Dr Hans Luyten is an assistant professor at the Department of Educational Organisation and Management, Faculty of Behavioural Sciences, University of Twente, Enschede, the Netherlands.

References

Beaton, A. E., Mullis, I. V. A., Martin, M. O., Gonzalez, E. J., Kelly, D. L. & Smith, T. A. (1996) *Mathematics achievement in the middle school years* (Chestnut Hill, TIMSS international study center, Boston College).

Beaton, A. E., Martin, M. O., Mullis, I. V. A., Gonzalez, E. J., Smith, T. A & Kelly, D. L. (1996) *Science achievement in the middle school years* (Chestnut Hill, TIMSS international study center, Boston College).

Blatchford, P., Goldstein, H., Martin, C. & Browne, W. (2002). A study of class size effects in English school reception year classes, *British Educational Research Journal,* 28(2), 169–185.

Brookover, W., Beady, C., Flood, P., Schweitzer, J. & Wisenbaker, J. (1979) *School social systems and student achievement: schools can make a difference* (New York, Praeger).

Cahan, S. & Davis, D. (1987) A between-grades-level approach to the investigation of the absolute effects of schooling on achievement, *American Educational Research Journal,* 24(1), 1–12.

Coe, R. & Fitz-Gibbon, C. (1998) School effectiveness research: criticisms and recommendations, *Oxford Review of Education,* 24(4), 421–438.

Cohen, J. (1988) *Statistical power analysis for the behavioral sciences, second edition.* (Hillsdale, Lawrence Erlbaum Associates, Publishers).

Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F. & York, R. (1966) *Equality of education opportunity* (Washington, DC, Government Printing Office).

Cooper, H., Nye, B., Charlton, K., Lindsay, J. & Greathouse, S. (1996) The effects of summer vacation of achievement test scores: a narrative and meta-analytic review, *Review of Educational Research,* 66, 227–268.

Edmonds, R. (1979) Effective schools for the urban poor. *Educational Leadership,* 37, 15–27.

Goldstein, H. (1997) Methods in school effectiveness research, *School Effectiveness and School Improvement*, 8, 369–395.

Giroux, H. A. (1983) *Theory and resistance in Education; a pedagogy for the opposition* (London, Heinemann).

Gonzalez, E. J. & Smith, T. A. (1997) *User Guide for the TIMSS International Database, primary and middle school years* (Chestnut Hill, TIMSS international study center, Boston College).

Heyns, B. (1978) *Summer learning and the effects of schooling* (New York, Academic Press).

Judd, C. M. & Kennedy, D. A. (1981) *Estimating the effects of social interventions* (New York, Cambridge University Press).

Lindahl, M. (2001) *Summer learning and the effect of schooling: evidence from Sweden* (Amsterdam, University of Amsterdam).

Martin, M. O., Mullis, I. V. S., Beaton, A. E., Gonzales, E. J., Smith, T. A. & Kelly, D. (1997) *Science achievement in the primary school years: IEA's Third International Mathematics and Science Study (TIMSS)* (Chestnut Hill: TIMSS international study center, Boston College).

Mortimore, P., Sammons, P., Stoll, L., Lewis, D. & Ecob, R. (1988) *School matters: the junior years* (Somerset, Open Books).

Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzales, E. J., Kelly, D. & Smith, T. A. (1997) *Mathematics achievement in the primary school years: IEA's Third International Mathematics and Science Study (TIMSS)* (Chestnut Hill: TIMSS international study center, Boston College).

Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D., Langford, I. & Lewis, T. (2000) *A user's guide to Mlwin* (London, Multilevel Models Project, Institute of Education, University of London).

Rossi, P. H., Freeman, H. E. & Lipsey, M. W. (2004). *Evaluation, a systematic approach, seventh edition* (Thousand Oaks/London/New Delhi, SAGE Publications).

Scheerens, J. & Bosker, R. J. (1997) *The foundations of educational effectiveness* (Oxford, Pergamon).

Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002) *Experimental and quasi-experimental designs for generalized causal inference* (Boston/New York, Houghton Mifflin Company).

Snijders, T. A. B. & Bosker, R. J. (1999) *Multilevel analysis, an introduction to basic and advanced multilevel modeling* (London/Thousand Oaks/New Delhi, SAGE Publications).

Ten Dam, G., Volman, M., Westerbeek, K., Wolfgram, P., Ledoux, G. & Peschar, J. (2003) *Social competentie langs de meetlat, Het bestrijden en voorkomen van onderwijsachterstand, Het evaluerenen en meten van sociale competentie* (Social competency assessed, Fighting and preventing educational disadvantage, Evaluating and measuring social competency) (Amsterdam, Universiteit van Amsterdam/Instituut voor de Lerarenopleiding).

Tymms, P., Merrell, C. & Henderson, B. (1997) The first year at school: a quantitative investigation of the attainment and progress of pupils, *Educational Research and Evaluation*, 3, 101–118.

Trochim, W. M. K. (1984) *Research design for program evaluation, the regression-discontinuity approach* (London, SAGE Publications).

Van der Wal, M. (2004) *Competencies to participate in life* (Groningen, Rijksuniversiteit Groningen).

Appendix: Curvilinear relations between date of birth and achievement?

Modelling the relation between date of birth and achievement is of crucial importance in the regression-discontinuity approach. If a linear function is estimated, while in reality the function is quadratic or cubic, the model will be misspecified and the regression-discontinuities may be biased. This appendix presents the findings of the analyses of the (curvi)linearity of the relationship between date of birth and achievement. The quadratic-function models (2b) and the cubic-function model (2c) are compared to the models with straightforward linear functions (model 2). The linear models are also presented in tables 4 and 5. None of the quadratic or cubic terms produced an effect that was statistically significant at the .05 level in a two-tailed test. All t-values (i.e. the ratios of the effects to the corresponding standard errors) are within the range of –1.96 to +1.96. Two t-values are outside the range of –1.65 to +1.65. Both effects relate to science achievement (i.e. the quadratic term in model 2b for Japan and the cubic term in model 2c for Greece). This corresponds to a significance level of .10 (two-tailed). If one applies a significance criterion this moderate one may expect to find three 'significant' effects merely by chance, as more than thirty models were fitted. The analyses indicate that a linear function provides an adequate description of the relationship between age and mathematics or science achievement. The most plausible explanation for this is the restricted age range (two years) of the students in this study. It seems likely that a curvilinear relationship can be established by extending the range of the analyses.

Appendix A. Curvilinear relations between date of birth and mathematics achievement

	Model	Intercept		Grade level		Date of birth		Date of birth quadratic		Date of birth cubic	
Cyprus	2	448.59	(3.53)	40.39	(5.91)	−30.40	(4.28)	–	–	–	–
	2a	447.63	(3.37)	40.29	(4.85)	−30.08	(3.52)	3.14	(3.39)	–	–
	2b	446.30	(3.67)	43.68	(6.12)	−22.89	(8.69)	1.99	(3.62)	−8.08	(8.92)
England	2	470.50	(4.43)	21.98	(4.69)	−35.23	(3.65)	–	–	–	–
	2a	471.21	(4.57)	22.02	(5.04)	−35.37	(3.69)	−2.23	(3.58)	–	–
	2b	473.42	(4.80)	16.54	(6.22)	−47.74	(9.00)	−.52	(3.75)	13.92	(9.23)
Greece	2	438.27	(4.38)	44.88	(5.32)	−22.85	(3.51)	–	–	–	–
	2a	438.02	(4.51)	44.88	(5.32)	−22.74	(3.54)	.79	(3.41)	–	–
	2b	439.69	(4.73)	40.56	(6.44)	−32.23	(8.71)	2.42	(3.67)	10.73	(9.01)
Iceland	2	420.95	(3.18)	39.15	(5.07)	−23.22	(4.38)	–	–	–	–
	2a	420.35	(3.56)	39.23	(5.05)	−23.03	(4.17)	1.74	(4.03)	–	–
	2b	421.37	(3.96)	36.74	(6.60)	−28.50	(10.20)	2.53	(4.25)	6.19	(10.55)
Japan	2	549.99	(2.30)	32.53	(3.77)	−26.47	(3.34)	–	–	–	–
	2a	551.06	(2.45)	32.56	(3.45)	−26.77	(2.93)	−3.35	(2.84)	–	–
	2b	551.14	(2.73)	32.35	(4.55)	−27.25	(7.23)	−3.29	(2.99)	.53	(7.47)
Norway	2	433.35	(3.44)	58.10	(5.05)	−22.47	(3.71)	–	–	–	–
	2a	434.60	(3.65)	58.11	(5.05)	−22.81	(3.73)	−3.74	(3.57)	–	–
	2b	436.46	(3.98)	53.56	(6.38)	−32.55	(9.14)	−2.46	(3.73)	10.81	(9.26)
Scotland	2	470.28	(3.85)	35.71	(4.71)	−28.90	(3.42)	–	–	–	–
	2a	469.57	(3.98)	35.61	(4.71)	−28.75	(3.43)	2.27	(3.29)	–	–
	2b	470.35	(4.22)	33.62	(8.46)	−32.97	(8.37)	2.96	(3.52)	4.67	(8.46)
Singapore	2	552.06	(4.72)	55.74	(5.63)	−18.27	(2.41)	–	–	–	–
	2a	553.47	(4.78)	55.64	(5.63)	−18.53	(2.42)	−4.12	(2.34)	–	–
	2b	552.47	(4.90)	57.96	(6.16)	−13.33	(6.07)	−4.71	(2.42)	−5.81	(6.22)

Only fixed effects are reported

The figures between brackets denote the standard errors

Appendix B. Curvilinear relations between date of birth and science achievement

	Model	Intercept		Grade level		Date of birth		Date of birth quadratic		Date of birth cubic	
Cyprus	2	432.49	(3.06)	28.96	(5.48)	−28.37	(4.13)	–	–	–	–
	2a	431.80	(2.97)	28.88	(4.34)	−28.14	(3.20)	2.26	(3.07)	–	–
	2b	431.76	(3.25)	28.97	(5.51)	−27.94	(7.89)	2.23	(3.28)	−.22	(8.10)
England	2	512.05	(4.97)	18.44	(5.46)	−35.37	(4.38)	–	–	–	–
	2a	512.13	(4.94)	18.45	(5.53)	−35.39	(4.13)	−.25	(4.00)	–	–
	2b	514.46	(5.21)	12.67	(6.87)	−48.39	(10.08)	1.56	(4.20)	14.63	(10.34)
Greece	2	456.52	(4.07)	31.03	(5.23)	−23.77	(3.68)	–	–	–	–
	2a	457.09	(4.22)	31.04	(4.89)	−24.02	(3.36)	−1.83	(3.24)	–	–
	2b	459.55	(4.43)	24.69	(5.98)	−38.01	(8.28)	.56	(3.49)	15.83	(8.56)
Iceland	2	448.12	(3.92)	41.45	(6.01)	−27.53	(4.92)	–	–	–	–
	2a	446.63	(4.36)	41.63	(6.06)	−27.05	(4.97)	4.28	(4.80)	–	–
	2b	447.27	(4.83)	40.08	(7.89)	−30.45	(12.15)	4.78	(5.06)	3.85	(12.56)
Japan	2	534.60	(2.00)	24.73	(3.36)	−27.15	(2.84)	–	–	–	–
	2a	536.15	(2.23)	24.77	(3.33)	−27.59	(2.72)	−4.85	(2.64)	–	–
	2b	536.48	(2.49)	23.94	(4.33)	−29.43	(6.71)	−4.59	(2.78)	2.08	(6.94)
Norway	2	465.51	(4.05)	47.11	(6.27)	−33.41	(4.55)	–	–	–	–
	2a	467.53	(4.31)	47.11	(6.26)	−33.98	(4.57)	−6.04	(4.38)	–	–
	2b	467.23	(4.74)	47.85	(7.88)	−32.41	(11.21)	−6.25	(4.58)	−1.75	(11.36)
Scotland	2	497.05	(5.03)	21.92	(5.50)	−32.38	(4.34)	–	–	–	–
	2a	495.57	(4.91)	21.75	(5.12)	−32.03	(3.87)	4.62	(3.71)	–	–
	2b	497.36	(5.16)	17.20	(6.53)	−41.69	(9.43)	6.20	(3.97)	10.69	(9.53)
Singapore	2	486.17	(4.62)	46.10	(5.29)	−14.16	(2.40)	–	–	–	–
	2a	487.27	(4.76)	46.02	(5.13)	−14.37	(2.31)	−3.22	(2.23)	–	–
	2b	487.26	(4.87)	46.03	(5.65)	−14.34	(5.79)	−3.22	(2.31)	−.03	(5.94)

Only fixed effects are reported
The figures between brackets denote the standard errors