

# Are classrooms equalizers or amplifiers of inequality? A genetically informative investigation of educational performance

Kim Stienstra<sup>1,\*</sup>, Antonie Knigge<sup>1</sup>, Ineke Maas<sup>1,2</sup>, Eveline L. de Zeeuw<sup>3</sup> and Dorret I. Boomsma<sup>3,4,5</sup>

<sup>1</sup>Department of Sociology/ICS, Utrecht University, Utrecht, The Netherlands

<sup>2</sup>Department of Sociology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

<sup>3</sup>Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

<sup>4</sup>Amsterdam Public Health Research Institute, Amsterdam, The Netherlands

<sup>5</sup>Amsterdam Reproduction & Development Research Institute, Amsterdam, The Netherlands

\*Corresponding author: Email: [i.k.stienstra@uu.nl](mailto:i.k.stienstra@uu.nl)

We investigate the influence of the classroom environment on educational performance and its dependency on parental socio-economic status (SES). The classroom environment can have a compensatory effect and decrease educational inequality, in which case the classroom context is more important for children originating from lower SES families. Alternatively, there can be an amplifying effect, in which case the classroom environment is more important for high-SES children. This would increase educational inequality. We investigate the two alternatives by applying a twin design to data from 4,216 twin pairs from the Netherlands Twin Register (birth cohorts 1991–2002). Some twin pairs share a classroom and other twins from the same pair are in different classrooms. We use this fact to decompose the variance in educational performance at the end of primary school into four components: genetic variance, classroom variance, shared environmental variance, and non-shared environmental variance. We find that of the total variance in educational performance, only a small part (2 per cent) can be attributed to differences between classrooms within schools. The influence of the classroom was larger when the level of parental SES was lower (up to 7.7 per cent) indicating a compensatory effect.

## Introduction

Do schools reduce social inequality in educational performance and serve as ‘the Great Equalizer’? Or do they reproduce or even amplify inequalities? Over the past century, these questions have been a core focus of social science scholars and central to the debate on educational policies and reforms (Coleman et al., 1966; Downey, von Hippel and Broh, 2004; Downey and Condron, 2016). There are diverging ideas on how children from different socio-economic status (SES) backgrounds benefit from their school environment. On the one hand, schools could amplify inequality. Even if low-SES and high-SES children attend the same school, various opportunities and practices in the classroom may favour the educational performance of high-SES children. For example, high-SES students could profit more from the materials and

lessons in effective classrooms because they enter the school with better academic preparation (Stanovich, 1986; Hanselman, 2018). On the other hand, schools could equalize inequality (i.e. a compensatory effect). Learning opportunities in schools and families overlap and could substitute for each other. An advantageous classroom environment would then benefit students from less resourceful family environments more (Chiu and Khoo, 2005; Hanselman, 2018).

Given the concerns about (increasing) educational inequality in many countries (OECD, 2019), it is important to understand to what extent the classroom context works as an equalizer and is part of the solution, or contributes to inequality and is part of the problem. Therefore, we ask: ‘To what extent are classroom effects contingent on SES background?’ Empirical studies did not provide a conclusive answer to this question

Received: July 2021; revised: September 2022; accepted: October 2022

© The Author(s) 2022. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

yet (e.g., Downey et al., 2004; Alexander, Entwisle and Olson, 2007; Downey and Condrón, 2016; Hanselman, 2018; von Hippel, Workman and Downey, 2018; Kyriakides, Creemers and Charalambous, 2019). The reasons for the mixed findings remain unclear. They do not seem to be dependent on students' age or country, for example.

Investigating classroom effects on educational performance is complex. The quality of the classroom environment is difficult to observe (Hanushek and Rivkin, 2006). For example, measures for teacher quality (e.g. experience, education, test scores) often have a small effect on educational performance (Rivkin, Hanushek and Kain, 2005). While teacher quality effects could be overstated, it could also be that measurable characteristics cover only a small part of the true variation in teacher quality (Rivkin et al., 2005). Another challenge is that classroom effects can be biased when alternative influences on educational performance are not sufficiently controlled for. High-SES parents more often choose high-quality schools for their children (Borghans, Golsteyn and Zölitz, 2015). Also, children's educational performance is genetically influenced (de Zeeuw et al., 2019), and the composition of classrooms concerning genetic predisposition may not be independent of the classroom environment. Children with a higher polygenic score for learning could be concentrated in higher quality classrooms (cf. Smith-Woolley et al., 2018). The classroom environment may capture influences of children's socio-economic background or genetic predisposition if those influences are not sufficiently considered.

Prior research tries to deal with such issues with value-added models. These models focus on students' learning gain instead of performance on a single occasion. Students' prior achievement is used to estimate the contribution ('added value') of the teacher or school to students' progress in achievement over time (OECD, 2018). Such value-added effects can be adjusted by student- and school-level covariates to further control for pre-existing differences between students, although it is debated to what extent such covariates should be included (Marks, 2021). Value-added models are often argued to provide a suitable way to estimate the contribution of classrooms or schools separate from alternative influences (see e.g. Wayne and Youngs, 2003), but they also come with disadvantages. One of these, which has only been limitedly acknowledged, is that value-added effects can be genetically influenced and may therefore provide biased estimates of the contribution of teachers and schools (Morris et al., 2018).

In this study, we take genetic influences into account by using a suitable method that is rarely applied in this research area. We estimate overall classroom effects and their dependency on parental SES by comparing

the educational performance of twins who attend the same or different classrooms. With this method, classroom influences are not directly measured but a latent factor, in addition to other latent factors capturing the contribution of genes, the shared environment (i.e. influences making twins raised in the same family more similar), and non-shared environment (i.e. influences making twins raised in the same family dissimilar) (e.g. Plomin et al., 2008). With this twin design, we can detect classroom effects that suffer less from insufficient measures and bias due to unobserved background influences. An advantage is that we do not only isolate classroom effects but also pinpoint the sources of alternative influences, that is, genetic, shared environmental, or non-shared environmental influences.

Two prior studies used this approach to estimate classroom influences in Australia and the United States (Byrne et al., 2010; Grasby et al., 2020).<sup>1</sup> They found that 0–9 per cent of the variance in literacy and numeracy could be accounted for by the classroom environment. We take the next step by investigating how classroom effects depend on parental SES. In this way, we investigate whether the classroom environment affects educational performance in an SES-neutral, equalizing, or amplifying way.

We analyse data on 4,216 twin pairs from the Netherlands Twin Register (NTR). We use the score on the nationwide standardized achievement test (i.e. *Cito*-test) to measure educational performance. Children take this test at the end of primary school around age 12. The test score, combined with the teacher's recommendation, determines enrolment in a secondary educational track. The *Cito*-score is thus important for children's educational careers. A key aspect of primary education in the Netherlands is that within-school tracking or academic streaming—a practice where classroom allocation is based on prior educational performance—is uncommon. This means that being in the same classroom does not depend on similarity in educational performance between twins, making the Dutch context well suited for investigating classroom effects with the twin method.

## Theoretical background

### Classroom influences

A classroom component derived from twin models is an omnibus measure for several interrelated ways through which the classroom environment could affect educational performance. Although we do not test the specific influences, we provide an overview of these possible ways.

The first route is via *teacher quality*. Teachers play a key role in students' educational performance, either directly (e.g. via instructions) or indirectly (e.g.

contributing to a safe and orderly classroom climate). How well teachers do their job is difficult to observe (Hanushek and Rivkin, 2006). Hence, most research is based on observable teacher characteristics, such as knowledge and experience (Parcel and Dufur, 2001; Hanushek and Rivkin, 2006). Studies show that teacher experience matters, but generally only in the first years of their employment (see Coenen et al., 2018). Teachers' cognitive skills are found to be positively associated with student performance although there are indications that it matters for students' math performance but not for reading (see Coenen et al., 2018). Lastly, teaching certificates and advanced degrees (i.e. a Master's or PhD degree relative to a Bachelor's degree) are generally not associated with students' performance (Wayne and Youngs, 2003; Hanushek and Rivkin, 2006; Coenen et al., 2018).

A second feature involves classroom *resources*, for example, equipment, smaller class sizes, and a lower student–teacher ratio. Smaller class sizes and a better student–teacher ratio are thought to reduce resource dilution effects: students have fewer classmates to ‘compete’ for teachers' time and attention (Parcel and Dufur, 2001). Furthermore, smaller classes may have fewer disciplinary problems and therefore more instructional time and greater opportunity to learn (see Blatchford and Russell, 2020). In empirical studies, indicators such as per-pupil expenditure, student–teacher ratio, and class size have been used. Per-pupil expenditure shows a consistent positive association with educational performance (Greenwald, Hedges and Laine, 1996), whereas for student–teacher ratio and class size, the results are mixed (Greenwald et al., 1996; Blatchford and Russell, 2020).

The third aspect is classroom *climate*. An academically oriented culture with high expectations signals certain standards about schoolwork and ideal student performance, which is thought to affect educational performance (Anderson, 1982). Also, a cohesive community with dense social ties could positively benefit students' performance (Parcel and Dufur, 2001; Dronkers and Robert, 2008). Such a community could affect students' sense of belonging, which relates to academic performance (Ma, 2003). Empirical studies show that classroom and school climate contributes to educational outcomes, with some evidence that it does not only influence short-term performance but that its effect persists for years (Thapa et al., 2013).

Lastly, *peers* in the classroom can play a role. One way is via direct interactions in learning. This does not only relate to figuring out an answer to an exercise, but also to processes such as developing vocabulary and obtaining knowledge that other students gained from museums, travelling, and so on (Kahlenberg, 2001; Hanushek et al., 2003). Additionally, peers influence

each other's motivation, aspirations, and attitudes towards education (Kahlenberg, 2001; Hanushek et al., 2003). Furthermore, peers affect the extent to which an environment is conducive to learning (Kahlenberg, 2001). If other students show disruptive behaviour, this affects classroom processes and lowers the learning opportunities for all students (Lazear, 2002). Conversely, highly motivated and skilled students contribute to a learning-oriented peer culture (Rumberger and Palardy, 2005). Numerous empirical studies show that peer influence is positively associated with students' educational outcomes (e.g. grades, standardized test scores), and other outcomes contributing to educational performance including motivational outcomes (e.g. goals, efficacy, interest) and positive behavioural styles (e.g. cooperative behaviour, self-confidence) (see Wentzel and Ramani, 2016). Given these mechanisms and prior empirical studies, we expect that the classroom context influences children's educational performance.

### The moderating role of parental SES

Children's educational performance is strongly affected by their family background. Children from a high-SES background perform on average better in school because of genetics, and the financial, cultural, and social resources high-SES parents can invest in children's educational success (Blau and Duncan, 1967; Bourdieu and Passeron, 1977).<sup>2</sup> We examine whether parental SES also affects the extent to which the classroom environment matters for educational performance.

There could be a *compensatory effect* in which case the educational performance of children from low-SES backgrounds depends more on their school environment. A reason to expect this is that learning opportunities within schools overlap with those within socio-economically advantaged families and can substitute for each other (Chiu and Khoo, 2005; Hanselman, 2018). A low-quality classroom environment may be less harmful to high-SES students because the fewer learning opportunities in such classes can be substituted by parental resources (e.g. providing tutoring), while low-SES parents cannot provide such compensation (Hanselman, 2018). In a similar vein, a higher quality classroom environment may be more beneficial for students from less advantaged family backgrounds. Low-SES students may be more susceptible to the supportive environment in advantaged classrooms because they are less likely to find supportive influences at home (Coleman et al., 1966; Rumberger and Palardy, 2005). For example, children in low-SES families receive less language stimulation at home as low-SES parents, in general, tend to expose their children to less vocabulary and grammar, read to

them less, and purchase fewer reading materials for their children (Bradley and Corwyn, 2002). These children may benefit more from the available books in the classroom and interactions with peers that contribute to developing vocabulary.

Also, the opposite can be expected: the classroom environment may be especially beneficial for high-SES students. This refers to an *amplification effect* and has also been referred to as multiplication and multiplicative accumulation (Erola and Kilpi-Jakonen, 2017) or cumulative advantage (DiPrete and Eirich, 2006). High-SES students may take greater advantage of the classroom environment because they enter this environment with better academic preparation (Hanselman, 2018). For example, high-SES students enter school with a stronger vocabulary. They may, therefore, benefit more from learning opportunities in school than low-SES students, because they may understand instructional materials better and learn more from particular lessons than low-SES students (Stanovich, 1986; Hanselman, 2018).

Another consideration is the cultural correspondence between the classroom and the home environment. From a cultural reproduction perspective (Bourdieu and Passeron, 1977), the high-status cultural signals of high-SES children (e.g. behaviours, tastes, and attitudes) are positively evaluated by teachers and these children may experience a greater sense of belonging in their class, leading to better educational performance (de Graaf, de Graaf and Kraaykamp, 2000). This could be especially the case in high-quality classrooms. In such classrooms, the culture is more ambitious and academically oriented, which coincides with high-SES parents' expectations and ambitions for educational success. This may further increase high-SES students' educational performance. For low-SES students, such a culture means a mismatch between their family and classroom experiences which may lead to negative self-perceptions and emotional distress, negatively affecting educational outcomes (Crosnoe, 2009). Additionally, the transmission of norms may be stronger if multiple actors socialize with students in the same way (Centola and Macy, 2007). Therefore, if the culture at home and in the classroom matches and both parents and teachers emphasize, for example, the importance of educational performance, students may be more likely to internalize this norm and behave accordingly.

Empirical tests of compensation and amplification effects based on a twin design have not been performed. Prior empirical studies investigated these effects through an interaction between parental SES and measured classroom (or school) characteristics, or by investigating the 'overall' contribution of classrooms or schools (e.g. with a value-added approach), all with mixed results. Studies with measured characteristics

focused on many different aspects of the classroom and found no interaction with SES (e.g. concerning teacher quality: Borman and Kimball, 2005; Akiba et al., 2007), compensatory effects (e.g. concerning student composition and climate, see e.g. Gustafsson, Nilsen and Hansen, 2018), and amplification effects (e.g. concerning resources, climate, and teacher quality; see Gustafsson et al., 2018; Atlay et al., 2019).

The results of studies that investigated the overall classroom effect are also inconclusive. For example, the value-added study by Kyriakides et al. (2019) shows that more effective classrooms and schools compensate for the socio-economic achievement gap. Additionally, the absence of schooling, for instance, during the summer break (e.g. Downey et al., 2004; Alexander et al., 2007) or COVID-19 pandemic (Engzell, Frey and Verhagen, 2021), increases the SES gap in educational outcomes, consistent with a compensatory effect of schooling. Yet, there are also comparable studies that challenge this conclusion. For example, Von Hippel, Workman and Downey (2018) find in their seasonal comparison study a small increase in SES inequality during the first summer and no change in inequality during the second summer. Additionally, Hanselman (2018) used a value-added approach and found no interplay with economic background, which suggests that classrooms and schools neither amplify nor compensate for socio-economic inequality but rather reflect existing inequality.

Since there are neither clear theoretical nor empirical arguments to favour either compensation or amplification, we consider two alternative hypotheses: *the classroom environment has a larger influence on educational performance if students' SES background is lower (H1; compensation)* and *the classroom environment has a larger influence on educational performance if students' SES background is higher (H2; amplification)*.

## Methods

### Data

We analyse data from the NTR, which was established in 1986 by recruiting twins and multiples a few weeks or months after birth. The NTR is still ongoing and registers around half of all newborn multiples in the Netherlands (Boomsma et al., 2002). For young twins, parents receive a survey at registration and when twins are 2, 3, 4/5, 7, 9/10, and 12 years old. After obtaining parental consent, twins' teachers receive surveys when the twins are 7, 9/10, and 12 years old. More details on the NTR are reported elsewhere (e.g. van Beijsterveldt et al., 2013; Ligthart et al., 2019).

We include twins from birth cohorts 1991–2002. For these cohorts, *Cito*-data are available for at least one

of the twins for 5,672 twin pairs (1,943 monozygotic [MZ] pairs and 3,729 dizygotic [DZ] pairs). The NTR determines twins' zygosity based on questionnaire items and on DNA or blood group polymorphism. The questionnaire allows for zygosity determination with 97 per cent accuracy (Ligthart et al., 2019).

For classroom placement, we rely on the mother's, father's, and teacher's reports from when the twins were 12 years old. Changing classrooms is uncommon in the Netherlands. The large majority of pupils share their classroom with the same children throughout primary school (Polderman et al., 2010). We initially rely on mother's report. If data were missing, we used father's report, and if both parents' reports included missing data, we used the teacher's report. Parents were asked which school situation is or was most applicable: (1) same class, (2) same school, parallel class, (3) same school, different class, (4) different school, (5) partly same, partly different. Class information from the teacher report was also measured when twins were around 12 years old, but in a less specific way as only a distinction between 'in the same class' and 'not in the same class' was made. We excluded those twin pairs for whom it was unclear whether they were in the same class or not, either due to missing data ( $N_{\text{pairs}} = 897$ , 15.8 per cent) or because they were partly in the same and partly in different classes ( $N_{\text{pairs}} = 333$ , 5.9 per cent), leading to a sample of 4,442 twin pairs. Additionally, we repeated the analyses after also excluding twins who went to different schools ( $N_{\text{pairs}} = 175$ , 3.1 per cent) and those who were in the same school but in different grades ( $N_{\text{pairs}} = 657$ , 11.6 per cent). The conclusions remain the same (see Appendix B in Supplementary Material).

Lastly, we excluded twin pairs with missing information on parental SES ( $N_{\text{pairs}} = 226$ ) as missingness on the moderator is not allowed. We used Full-Information Maximum Likelihood (FIML) estimation (Arbuckle, 1996) to handle missing values for all other variables. Our final sample consists of 4,216 twin pairs (880 MZ pairs in the same class, 596 MZ pairs in different classes, 1,444 DZ pairs in the same class, and 1,296 DZ pairs in different classes).

## Measurements

Our dependent variable, *educational performance*, is measured by the student's score on the *Cito*-test. These were initially obtained via teacher reports and later also via parents' and children's reports. *Cito*-scores from the different sources are highly correlated: the correlation between *Cito*-scores reported by parents and twins is 0.975 and between teachers and twins is 0.932 (van Beijsterveldt et al., 2013). The *Cito*-test is a nationwide standardized educational achievement test that is taken at the end of primary education (around

age 12). It consists of multiple-choice items on Dutch language, mathematics, study skills, and world orientation (e.g. geography, biology, and history). The domains are combined into a total score using Item Response Theory, and this score is standardized on a scale from 501 to 550. Because the sub-domain world orientation is not mandatory, this is not included in the calculation of the total score. The national average is a score of 535, with a standard deviation (SD) of 10. Our sample has a somewhat higher average and lower SD (see Table 1). Means and variances of the *Cito*-scores for twin 1 (typically the firstborn twin) and twin 2 are not statistically significantly different.

We measure parental SES by *parental education*, which is the most stable and important indicator of SES when predicting children's educational performance (Sirin, 2005). We use the information on mother's and father's highest educational level from the parents' survey when the twins were around 10 years old. When mother's or father's education at age 10 was missing, we used information from the survey for younger twins (ages 7, 3, and 1). The original variable from the parent survey for 10- and 7-year-old twins consisted of 13 categories ranging from elementary school to post-graduate degree/PhD. The variable measured at twin ages 3 and 1 consisted of five categories ranging from elementary school to scientific education. We converted these categories into scores on the International Standard Level of Education (ISLED) scale (Schröder and Ganzeboom, 2014). ISLED is a well-validated continuous education measure ranging from 0 to 100, which allows for comparison across surveys and countries. We coded the categories into the highest level of finished education using the online appendix (ISLED, 2014) and used the average ISLED score when multiple values were applicable (e.g. for the category combining higher general secondary education 'havo' and pre-university education 'VWO'). An overview of the coding is presented in Table A1 (Appendix A in Supplementary Material).

In all analyses, we control for *sex* (girls = 0, boys = 1) and *age* (in years) when the *Cito*-test was taken. While most children take the test when they are 12 years old, there is variation in age (see Table 1) which we want to correct for.

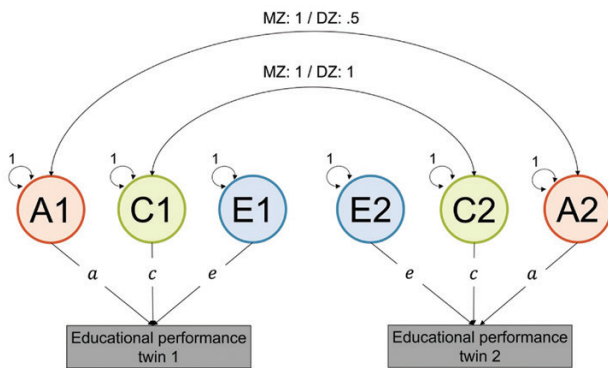
## Twin design

The classical twin design (Figure 1) decomposes the variance in educational performance into variance due to additive genetic influences (*A*), common or shared environmental influences (*C*), and non-shared or unique environmental influences including measurement error (*E*) (Franić et al., 2012). Twin data enable disentangling these variance components, as twins differ in genetic relatedness (MZ twins share almost 100 per cent of their

**Table 1.** Descriptive statistics for MZ twins in the same classroom, DZ twins in the same classroom, MZ twins in different classrooms, and DZ twins in different classrooms

Variable	Same classroom					Different classrooms				
	Mean	SD	Min.	Max.	N	Mean	SD	Min.	Max.	N
MZ twins										
Twin specific										
Cito twin 1	538.07	8.27	508	550	872	538.42	8.33	509	550	562
Cito twin 2	537.57	8.49	505	550	866	538.36	8.20	510	550	552
Boy twin 1	0.44	-	0	1	880	0.49	-	0	1	595
Boy twin 2	0.44	-	0	1	880	0.49	-	0	1	595
Twin pair										
Education parents	63.95	17.64	22.98	92.63	880	68.72	17.33	22.98	92.63	596
Age	12.26	0.49	11	14	740	12.27	0.52	11	14	472
$r_{cito1,cito2}$	.80					.77				
DZ twins										
Twin specific										
Cito twin 1	537.82	8.83	503	550	1,415	537.78	8.46	510	550	1,132
Cito twin 2	537.74	8.60	501	550	1,416	537.73	8.34	509	550	1,110
Boy twin 1	0.49	-	0	1	1,444	0.52	-	0	1	1,294
Boy twin 2	0.47	-	0	1	1,443	0.53	-	0	1	1,296
Twin pair										
Education parents	64.80	17.29	22.98	92.63	1,445	67.77	17.55	22.98	92.63	1,296
Age	12.25	0.47	10	14	1,181	12.28	0.51	11	14	936
$r_{cito1,cito2}$	0.47					0.44				

Note: For twin-specific variables, the N refers to number of individuals. For twin pair variables, the N refers to number of pairs.



**Figure 1.** The Classical Twin Design (CTD). The CTD is fitted to data from monozygotic (MZ) and dizygotic (DZ) twins. Latent variables (circles) represent genetic (A), shared environmental (C), and non-shared environmental (E) factors. Their influence on educational performance (observed variables: squares) is given by path coefficients *a*, *c*, and *e*.

genes at conception, DZ twin pairs share on average 50 per cent of their segregating genes) and MZ and DZ twin pairs are assumed to share their environment to the same

extent. Hence, the covariance in educational performance between twin 1 and twin 2 is  $Cov_{MZ} = a^2 + c^2$  for MZ twins and  $Cov_{DZ} = 0.5a^2 + c^2$  for DZ twins. A larger similarity in performance for MZ twins than DZ twins is consistent with a hypothesis that genetic influences are of importance. If MZ twins are not alike, the source of this dissimilarity is by definition the non-shared environment.

We estimate a fourth component capturing classroom influences (*CL*) using measured information on whether twins attend the same classroom in primary school or not. When twins share their classroom, classroom influences increase twin similarity in educational performance. Classroom influences make twins dissimilar if they are in different classrooms. The difference between twin correlations of twins in the same and different classrooms is the basis to estimate classroom influences (Byrne et al., 2010). As can be seen in Figure 2, the covariances in educational performance between twin 1 and twin 2 for the groups MZ twins in the same classroom (MZSC), MZ twins in different classrooms (MZDC), DZ twins in the same classroom (DZSC), and DZ twins in different classrooms (DZDC) are

$$\text{Cov}_{\text{MZSC}} = a^2 + c^2 + cl^2$$

$$\text{Cov}_{\text{MZDC}} = a^2 + c^2$$

$$\text{Cov}_{\text{DZSC}} = 0.5a^2 + c^2 + cl^2$$

$$\text{Cov}_{\text{DZDC}} = 0.5a^2 + c^2$$

Estimating classroom variance changes the interpretation of  $C$  and  $E$ . These components now refer to shared environmental and non-shared environmental influences not related to the classroom. The classroom component mainly captures within-school differences, because most twins attend the same school. Since between-school differences in the classroom environment are very limited in our case, this may lead to an underestimation of the classroom effect compared to regular multilevel studies of children in classes. This affects the interpretation of the results but does not prohibit testing our hypotheses.

We use multi-group Structural Equation Modelling to decompose the variance in educational performance into  $A$ ,  $C$ ,  $CL$ , and  $E$ . The model includes the measured educational outcome for twin 1 and twin 2 for

the groups MZSC, MZDC, DZSC, and DZDC. Latent factors  $A$ ,  $C$ ,  $CL$ , and  $E$  are set to a variance of one to identify the model. Path coefficients  $a$ ,  $c$ ,  $cl$ , and  $e$  represent the effects of the standardized latent factors on the observed outcome. The total variance is given by summing all squared path estimates:

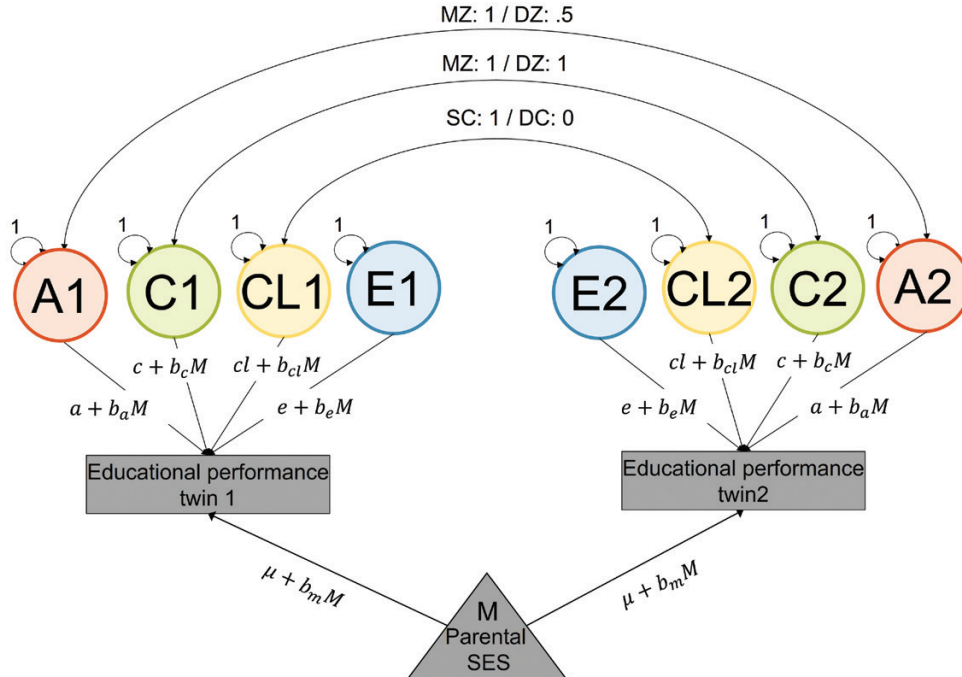
$$V_{\text{total}} = V_A + V_C + V_{CL} + V_E = a^2 + c^2 + cl^2 + e^2.$$

Variance components can be standardized. For example, the proportion of the genetic variance component to the total variance in educational performance, which is called heritability, is given by

$$\frac{a^2}{a^2 + c^2 + cl^2 + e^2} = \frac{V_A}{V_{\text{total}}}.$$

The contribution of genetic and environmental factors can depend on a moderator, in our case, parental SES (see Figure 2). For example,  $cl$  becomes  $cl + b_{cl}M$ , where  $M$  is the level of parental SES. The total variance in this moderation model becomes

$$V_{\text{educ} | M} = (a + b_a M)^2 + (c + b_c M)^2 + (cl + b_{cl} M)^2 + (e + b_e M)^2.$$



**Figure 2.** ACE moderation model extended with a classroom factor ( $CL$ ). Latent variables represent genetic ( $A$ ), shared environment ( $C$ ), classroom ( $CL$ ), and non-shared environment ( $E$ ) factors, with the corresponding path coefficients  $a$ ,  $c$ ,  $cl$ , and  $e$ . Estimated from data of monozygotic (MZ) and dizygotic (DZ) twins who are in the same classroom (SC) or different classrooms (DC). The correlation between CL1 and CL2 equals 1 if twins share a classroom, otherwise it is zero. Moderator  $M$  is a measure of parental SES and is also included in the model as fixed effect (triangle). The model also includes fixed effects of age and sex (not shown to avoid clutter).

## Analytical strategy

We fitted a series of models in Mplus using FIML estimation. In all models, we control for age and sex. We do not take nesting into classes and/or schools into account, partly because of a practical reason as for 60 per cent of the twins we do not have a school identifier, and partly because the number of twin pairs in the dataset per school is low (on average 1.5 for those with a school identifier).

We decomposed the variance in educational performance with and without a classroom component. This shows how the classroom component is captured by  $C$  and  $E$  if it remains unmodelled. Then, we included our measure for parental SES, which is expected to have a main effect on educational performance and explain part of the total variance. Because parental SES is measured at the family level and always shared between twins, it can only explain shared environmental variance (Turkheimer et al., 2005). In our final model, we allowed the paths to be moderated by parental SES in a continuous, linear gene-environment moderation model (Purcell, 2002). With this model, we tested whether there is a compensatory ( $H1$ ) or amplifying ( $H2$ ) effect. If compensation takes place, classroom-level variance would be smaller in higher educated families (i.e.  $b_{cl}$  is negative), whereas in the case of amplification, classroom-level variance would be larger in higher educated families (i.e.  $b_{cl}$  is positive). Given that compensation and multiplication could occur simultaneously, a negative interaction effect means that compensation is stronger than possible amplification effects (and vice versa). Lastly, we elaborated on how selection into school classes may bias our results.

Before fitting the twin models, we checked whether the means and variances of educational performance were equal for twin 1 and twin 2 (first and second born), MZ and DZ twins, and twins in the same and different classrooms. Therefore, we fitted a saturated model that describes the data with no free parameters left (i.e. no constraints on means and variances) and compared this with models that included the constraints using likelihood-ratio tests. The assumptions of equal means and variances are met.

## Results

### Decomposition of educational performance

The decomposition of educational performance into the  $A$ ,  $C$ ,  $CL$ , and  $E$  components is presented in Model 2 of Table 2. The total variance in educational performance is  $72.001 (= 6.989^2 + 2.699^2 + 1.241^2 + 3.785^2)$ . The classroom variance is  $V_{CL} = 1.241^2 = 1.540$ . While for classroom variance, the path loading of 1.241 is significant ( $p = 0.022$ ), the variance of 1.540 is not ( $p$

$= 0.252$ ). This may be because the power for estimating a variance component is smaller than for a path loading. Standardizing this variance component shows that individual differences in educational performance are for 2.1 per cent related to differences in classroom environment within schools ( $V_{CL}/V_{total}$ ). Additionally, there is statistically significant shared environmental variance ( $V_C = 2.699^2 = 7.287$ ) and non-shared environmental variance ( $V_E = 3.785^2 = 14.325$ ), making up 10.1 per cent and 21.0 per cent of the total variance in educational performance, respectively. The largest source of variance in educational performance is genetic ( $V_A = 6.989^2 = 48.849$ ). This leads to a heritability estimate of 0.678, meaning that 67.8 per cent of the individual differences in educational performance are related to genetic differences.

If classroom variance is unmodelled, it is captured in the  $C$  and  $E$ , because part of the twin pairs share the same classroom while the other part does not. Model 1 (Table 2) shows that if the classroom variance is not modelled, the shared environmental variance is  $V_C = 2.880^2 = 8.294$  and the non-shared environmental variance is  $V_E = 3.862^2 = 14.913$ . Comparing these shared and non-shared environmental variances with those in Model 2 ( $V_C = 7.287$  and  $V_E = 14.325$ , respectively) shows that the classroom variance is to a larger extent captured in the  $C$ -component. This is as expected given that the twins in our sample were more often in the same than in different classrooms.

We can use the difference in the  $C$ -components from Models 1 and 2 to calculate the extent to which shared environmental influences can be attributed to shared classroom experiences. Around 11.5 per cent of the differences in educational performance can be attributed to shared environmental influences if classroom effects are not considered and this is 1.4 percentage points lower (10.1 per cent) if we take classroom effects into account. This shows that  $1.4/11.5 = 12$  per cent of the shared environmental influences can be attributed to the same classroom experiences. One could do the same for the non-shared environment ( $E$ ) and different classroom experiences, but this is not very informative given that  $E$  not only includes different environmental influences but also measurement errors.

### The role of parental SES

Model 3 (Table 2) includes the main effect of parental education, our parental SES measure, which has a positive significant effect on educational performance. Each unit increase in parental education is associated with 0.154 point increase in *Cito*-score ( $b = 0.154$ ,  $p < 0.001$ ). The standardized effect (not shown in Table 2) is 0.319, meaning that an SD increase in parental education is associated with 0.319 SD increase in educational performance. Parental education entirely



**Table 2.** Results of twin models for educational performance for MZ twins in the same classroom ( $N_{\text{pairs}} = 880$ ), DZ twins in the same classroom ( $N_{\text{pairs}} = 1,444$ ), MZ twins in different classrooms ( $N_{\text{pairs}} = 596$ ), and DZ twins in different classrooms ( $N_{\text{pairs}} = 1,296$ )

	Model 1		Model 2		Model 3		Model 4	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Intercept	551.703***	(3.387)	551.771***	(3.388)	535.201***	(3.255)	535.719***	(3.209)
Parental education					0.154***	(0.006)	0.158***	(0.007)
Path								
<i>a</i>	6.987***	(0.206)	6.989***	(0.206)	6.995***	(0.117)	9.212***	(0.447)
<i>c</i>	2.880***	(0.441)	2.699***	(0.496)	0.000	(0.000)	0.000	(0.000)
<i>cl</i>			1.241*	(0.542)	1.253*	(0.507)	3.428***	(0.916)
<i>e</i>	3.862***	(0.099)	3.785***	(0.121)	3.784***	(0.110)	4.884***	(0.397)
<i>b<sub>a</sub></i>							-0.034***	(0.006)
<i>b<sub>c</sub></i>							0.000	(0.000)
<i>b<sub>cl</sub></i>							-0.029**	(0.010)
<i>b<sub>e</sub></i>							-0.017**	(0.006)
Variance								
$V_A$	48.812***	(2.877)	48.849***	(2.884)	48.928***	(1.630)	a	
$V_C$	8.294**	(2.538)	7.287**	(2.678)	0.000	(0.000)	a	
$V_{CL}$			1.540	(1.345)	1.570	(1.270)	a	
$V_E$	14.913***	(0.767)	14.325***	(0.914)	14.320***	(0.830)	a	
$V_{\text{total}}$	72.019***	(1.371)	72.001***	(1.370)	64.819***	(1.236)	a	
Model fit								
# free parameters	30		31		40		44	
LL	-35,692.259		-35,691.177		-53,444.695		-53,380.945	
AIC	71,444.517		71,444.353		106,967.390		106,849.889	

Note: Estimates are unstandardized, controlled for age and sex in all models. SE = standard error,  $V_A$  = genetic variance,  $V_C$  = shared environmental variance,  $V_{CL}$  = classroom variance,  $V_E$  = non-shared environmental variance,  $V_{\text{total}}$  = total variance,  $LL$  = loglikelihood, AIC = Akaike Information Criterion. The variances are based on the squared path estimates *a*, *c*, *cl*, *e* (e.g.  $a^2 = 6.987^2 = 48.812$ ).

\*Not applicable, because in the moderation model the size of a variance component depends on the level of parental education.

\* $p < 0.05$ .

\*\* $p < 0.01$ .

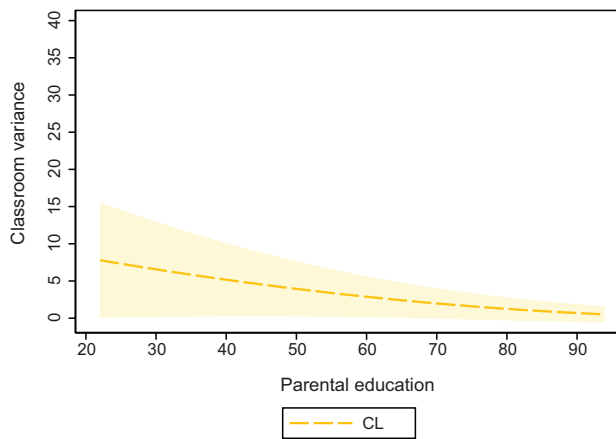
\*\*\* $p < 0.001$  (two-tailed tests).

explains the (non-classroom) shared environmental variance.<sup>3</sup>

In Model 4 (Table 2), we allow the variance components to be moderated by parental education. These interaction effects are indicated by  $b_a$ ,  $b_c$ ,  $b_{cl}$ , and  $b_e$ . If there is a compensatory effect (H1), classroom influences would become smaller with increasing parental education (i.e.  $b_{cl}$  is negative). In the case of an amplification effect (H2), we expect to observe the opposite (i.e.  $b_{cl}$  is positive). Our findings point towards a compensatory effect, as there is a negative statistically significant moderation of the classroom effect ( $b_{cl} = -0.029$ ,  $p = 0.006$ ). The classroom variance in this moderation model is computed by  $V_{CL|M} = (cl + b_{cl}M)^2$ . For children with the lowest educated parents (i.e. primary education—ISLED 22.98), classroom variance is

$V_{CL} = (3.428 + (-0.029 \times 22.98))^2 = 7.626$ . Classroom variance decreases and almost approaches zero ( $V_{CL} = 0.550$ ) for the highest educated parents (i.e. postdoctoral education—ISLED 92.63) (see Figure 3).

Classroom variance reduces with increasing parental education, but so does the total variance in educational performance (see Figure 4a). To investigate whether the compensation pattern appears because of the decreasing total amount of variance, we also look at the standardized results (see Figure 4b). The standardized results also show a pattern of compensation. We previously showed that for children with the lowest educated parents, the estimated classroom variance is 7.626. Dividing this by the total variance in educational performance among these children (98.893) shows that 7.7 per cent of the variance in educational performance



**Figure 3.** Unstandardized classroom variance (CL) in educational performance moderated by parental education (ISLED) including 95% confidence interval. Results based on Table 2 Model 4.

can be attributed to the classroom context. For the average parental education (higher and pre-scientific secondary school), 3.5 per cent can be attributed to the classroom, and for children with the highest educated parents (postdoctoral education) 1.2 per cent.

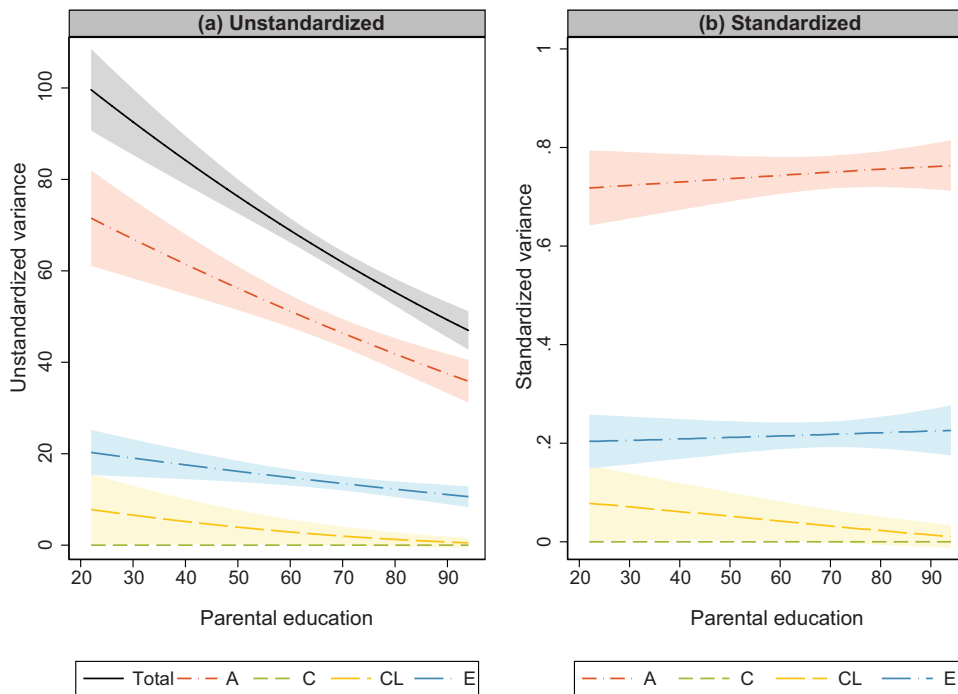
Although we did not hypothesize on the moderation of other variance components, results show statistically significant negative moderations of unstandardized genetic influences ( $b_a = -0.034$ ,  $p < 0.001$ ) and non-shared environmental influences ( $b_e = -0.017$ ,  $p = 0.002$ ). Yet, these largely disappear when we look at the standardized components (see Figure 4b). With increasing parental education, relative genetic variance and non-shared environmental variance increase a little bit but are largely stable around 0.75 and 0.20, respectively. Shared environmental variance is not dependent on parental education: it is entirely explained by parental education for all levels of parental education.

### Selection into classrooms

The Dutch Association of Parents of Multiples (NVOM) advises that the decision to allocate twins to the same versus different classrooms should be made on an individual basis by teachers and parents together (NVOM, 2019). To what extent schools follow this advice or have a different policy is not entirely clear. International studies show that official school policy to separate twins is exceptional. For example, a survey in the United Kingdom showed that only 1 per cent of the schools had an official policy on the education of twins (Preedy, 1999). Precise information on the Dutch context is lacking, although the NVOM (2019) reports that 9 per cent of the parents that participated in their survey indicate that only the school decides on the classroom allocation.

There may be, however, other selection effects. One of them relates to zygosity. An assumption underlying the classical twin design is that environmental influences are shared to the same extent by MZ and DZ twins. Under this assumption, greater similarity in educational performance among MZ twins compared to DZ twins can be attributed to MZ twins' greater genetic similarity. However, if MZ twins are more often in the same classroom than DZ twins, greater similarity in MZ twins' performance is due to both greater genetic similarity and greater classroom similarity. In our sample, MZ twins are somewhat more often in the same classroom than DZ twins (respectively, 59.6% vs. 52.7%,  $\chi^2 = 18.568$ ,  $df = 1$ ,  $p < 0.001$ ). If classroom sharing is unmodelled, this would lead to an overestimation of genetic influences (Grasby et al., 2020). Since we explicitly model classroom sharing, a bias is avoided in our study. Greater similarity among MZ twins is no longer an unobserved mixture of greater genetic similarity and being more often in the same classroom because incorporating classroom information allows for separating these sources.

While selection based on zygosity is not problematic for estimating classroom influences, selection based on other characteristics may be a reason to allocate twins to the same class or different classes. This is only problematic if such characteristics also affect our dependent variable: educational performance at age 12. Studies suggest that there are mainly two characteristics that affect the decision to place twins in separate classrooms: ability and behavioural problems (Jones and De Gioia, 2010). As these characteristics may also affect educational performance, the estimated classroom component could be inflated. Greater twin similarity in educational performance for twins in the same class would then reflect similarity in prior ability or (the lack of) behavioural problems, rather than being solely a consequence of exposure to the same classroom context. We do not expect such selection processes to play a major role in the Dutch context. In Dutch primary education, there is no ability tracking. Neither parents nor schools can choose a high-performing classroom for one twin and a low-performing classroom for the other. Moreover, prior twin studies investigating classroom effects in countries where within-school tracking is common indicate that class allocation based on prior performance did not affect the conclusions on classroom effects (Byrne et al., 2010; Grasby et al., 2020). Concerning behavioural problems, studies using Dutch twin samples show that the association between problem behaviour and classroom allocation, if any, is small and that it does not affect educational performance in the long run (i.e. performance at age 12) (van Leeuwen et al., 2005; Polderman et al., 2010). Thus, the Dutch



**Figure 4.** Decomposition of the (a) unstandardized and (b) standardized variance in educational performance moderated by parental education (ISLED). Genetic (A), shared environmental (C), classroom (CL), and non-shared environmental (E) variance. Results based on Table 2 Model 4.

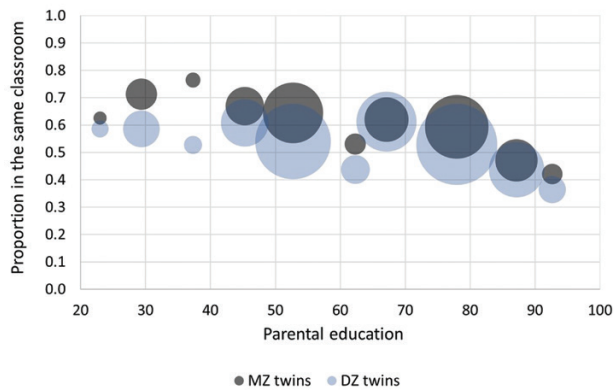
educational context and prior empirical evidence suggest that it is unlikely that originally dissimilar twins end up in different classes. Even if it would occur, the effect on the (over)estimation of classroom effects will likely be minimal.

Lastly, if the decision to place twins in different classrooms is selective (because of differences in performance or behavioural problems), it would influence the interaction that we are interested in if the bias depends on parental SES. Literature suggests that low-SES parents tend to be more ‘twinship oriented’ (i.e. treating twins more similar), while high-SES parents tend to be more ‘differentiation oriented’ (i.e. treating twins more dissimilar, emphasizing individuality), which may be especially pronounced for MZ twins (Robin and Casati, 1994). If so, our reported negative interaction effect between parental SES and influence of the classroom would be an underestimation, because the classroom effect would be overestimated for high-SES parents but not so much for low-SES parents. We do find that twins of lower SES parents are somewhat more often in the same class ( $\chi^2 = 78.407$ ,  $df = 9$ ,  $p < 0.001$ ) (see Figure 5). We do not expect the underestimation of the interaction to be large, because the SES difference in class allocation is small and our results indicate that high-SES parents are not more differentiation oriented than low-SES parents. Environmental

influences that make twins more dissimilar, such as parents treating twins differently, are captured in the non-shared environment component (E). If high-SES parents would indeed be much more differentiation oriented, one would expect E to be larger in high-SES families, which is not the case (see Figure 4a).

## Conclusion and discussion

In many countries, educational inequality based on socio-economic background is of great concern, leading researchers, policy-makers, and educational professionals to question how to counter this. Classrooms are important contexts in children’s lives, but it is unclear whether these are part of the problem by amplifying educational inequality or part of the solution by reducing educational inequality. We used a novel approach based on Dutch twin data and showed that the role of classrooms in explaining educational performance of primary school pupils is relatively minor when alternative sources (including genetic and non-shared environmental influences) are considered. However, classroom influences are stronger for children from lower SES families. This suggests that classrooms compensate for influences of family background and thus serve as equalizers rather than amplifiers of educational inequality.



**Figure 5.** Proportion of MZ and DZ twins in the same class by parental education (ISLED). The size reflects the number of observations.

We find that on average 2.1 per cent of the variance in educational performance of Dutch primary school pupils is attributable to classroom influences. This falls within the range (i.e. 0–9 per cent) that has been previously found in twin studies (Byrne et al., 2010; Grasby et al., 2020). There are different ways to interpret this result. On the one hand, an explained variance of 2.1 per cent indicates that only a small part of *all* differences in educational performance can be accounted for by the classroom environment. Given the general view that teachers and other classroom aspects are important for children’s educational outcomes, this may be somewhat surprising. It is important to be aware that our classroom effect shows the extent to which the classroom environment explains differences in performance, not the effect on average performance. For example, a good teacher may increase student performance but if most classes have good teachers, it would not explain much of the individual differences in performance. Yet, one would not conclude that good teachers are not important.

On the other hand, the classroom effect may also be interpreted as being more substantial if we relate it to the environmental part of the variance instead of the overall variance. Genetic variance is a major source of differences in educational performance. In our case, genetic differences account for 68 per cent of the total variance, meaning that only 32 per cent result from environmental factors. Compared to these environmental differences, the share of the variance that is explained by classrooms is more sizeable. Of all the environmental influences that children from the same family share (e.g. parents, neighbourhood, school), 12 per cent can be attributed to shared classroom experiences. Moreover, classrooms are more important for some children than for others, as we discuss next.

Classroom effects are dependent on family background. They are larger the lower educated the parents are, explaining up to 7.7 per cent of the total variance in educational performance for children from the lowest educated families. This suggests a compensatory effect and is consistent with the argument that children from low-SES families could substitute school resources for family resources. Because these children are less likely to find supportive influences at home, for example, they may be more susceptible to a supportive school environment (Coleman et al., 1966; Rumberger and Palardy, 2005). This does not mean that amplification effects do not take place at all. Favourable classroom environments could have a compensating effect for low-SES children while simultaneously having an amplifying effect for high-SES children. If this would occur to the same extent, we would not find any differences in the classroom component by SES. Since we find larger classroom effects for low-SES children, we can conclude that the compensatory effects of classrooms are stronger than possible amplification effects.

While we did not hypothesize on how other sources of variance depend on SES background, we find less genetic and non-shared environmental variance in higher SES families. Gene–environment interaction in educational outcomes is often studied from a bio-ecological framework, where the genetic potential for high educational performance is thought to be actualized in more advantaged environments such as high-SES families (Scarr-Salapatek, 1971; Bronfenbrenner and Ceci, 1994). However, we find that genetic variance is higher in low-SES families. This can be interpreted as genetic influences associated with lower educational performance having more detrimental effects on children from more disadvantaged socio-economic backgrounds (see de Zeeuw et al., 2019). When we consider that the total variance in educational performance decreases by standardizing the results, we do not find a moderation of genetic variance (nor non-shared environmental variance) by parental SES.

Although twin data allow us to contribute to prior studies by investigating classroom effects and their dependency on parental SES in a novel way, it does not solve all issues. In particular, the (moderating) effect of measured parental SES should not be interpreted as causal. Parents’ genetic potential influences parents’ SES, and they transmit part of this genetic potential to their children, which in turn influences the children’s school performance. If these genes that parents and their children share are unmodelled, the association between parental SES and children’s education is likely genetically confounded and thus does not only reflect environmental effects (e.g. effects of resources and behaviours shaped by parental SES) (Hart, Little and van Bergen, 2021). This issue is not specific to our twin analyses but applies to most analyses

involving parental SES and children's outcomes (except those with a causal design such as an instrumental variable or difference-in-difference approach). For our conclusions, this means that the observed compensation of the impact of family background by classrooms does not pertain only to environmental (dis)advantages passed on by families, but possibly also to the transmission of genetic (dis)advantages. This is not problematic if one is interested in the stratification of educational outcomes by family background in a broader sense, capturing all types of (dis)advantages that are associated with it. If one is interested in separating environmental and genetic intergenerational transmission, one would need other designs such as the children-of-twin design or measured genotype design.

Using twin data also comes with some complexities of its own. A concern may be that twins form a special group and that conclusions based on twin data are not generalizable to the general population. Twins, especially identical ones, may be raised in more similar environments and treated more similarly than fraternal twins or non-twins. This could affect estimations if differential treatment is related to the outcome (i.e. educational performance). We do not expect this to be the case. Prior studies show only limited evidence for a violation of the equal environment assumption and if a violation occurred it did not affect the genetic and environmental effects on educational outcomes (Mönkediek et al., 2020). Also, in a population-based study on educational performance in the Netherlands, twin-based estimates of *Cito*-scores were not biased (Schwabe, Janss and van den Berg, 2017). These findings suggest that our results are generalizable to the general population of Dutch primary school pupils.

Nevertheless, relying on twin data may lead to conservative estimates of classroom effects. Twins almost always attend the same school, meaning that we capture only within-school differences between classrooms. Some classroom characteristics (e.g. SES composition, school resources) cluster within schools, implying that the differences between classrooms in our study are smaller than the differences between classrooms in general. The classroom effect that we find, therefore, reflects aspects that vary between classrooms within a school (e.g. teacher quality, climate) rather than aspects that vary mostly between schools (e.g. resources, student composition). Future research could include measured classroom characteristics to investigate which specific characteristics contribute to classroom variance. Another reason why our estimate of classroom effects is conservative relates to non-shared environmental influences. The non-shared environment makes up around 20 per cent of the total variance in educational performance after distinguishing the classroom component. Part of this non-shared

environmental variance may still be related to what happens in classrooms. Even if twins are in the same classroom, they may perceive their classroom environment differently or teachers could treat them differently, which is captured in the non-shared environment component. Altogether, this means that our classroom estimate should be seen as a lower bound. Future work using complementary approaches (e.g. children-of-twins models, parental genotype design, including individual-specific measures of the class environment) could provide more insight into the interrelated influences of genes, families, and classrooms.

In conclusion, we find that classroom influences depend on family SES in Dutch primary schools. There is more classroom variance with lower levels of parental education, suggesting that children from lower SES families benefit more from a high-quality classroom than children from high-SES families. However, we cannot conclude that classroom environments are great equalizers because the role of classrooms in explaining differences in educational performance is relatively small. Moreover, it can be expected that especially children from high-SES families are in high-quality classes given socio-economic selection into schools (Borghans et al., 2015). High-SES children are more likely to be in high-quality classes while the added value of such an environment is relatively little for them. Contrarily, low-SES children, for whom a high-quality class environment could make more of a difference, are likely less often exposed to this. Our results are thus indicative of a compensatory effect and show the modest potential of the classroom environment to reduce educational inequalities. Whether this potential is actually realized depends on the quality of the classroom environment that low-SES children are exposed to.

## Notes

- 1 A few other twin studies also provide information on twins in the same and different classes but do not explicitly test for classroom variance (e.g. Kovas et al., 2007; Eifler et al., 2019).
- 2 Because parental genes can influence both parental SES and children's educational performance (via children's genes), the association between parental SES and children's educational performance may be partly spurious. We come back to this issue and its implications in the Conclusion and Discussion section.
- 3 This also leads to decreased model fit which would be less if we drop the C-component. We did not do so because C-variance may still be important for part of the parental education scale when we estimate the moderation model.

## Supplementary Data

Supplementary data are available at *ESR* online.

## Acknowledgements

We are grateful to the twin families and the teachers for their participation. Earlier versions of this work were presented at the ISA RC28 spring meeting (Frankfurt, 2019), Dag van de Sociologie (Amsterdam, 2019), Social Stratification Research Seminar (Amsterdam, 2019), Migration and Social Stratification Seminar (Utrecht, 2019), and ICS spring day (Nijmegen, 2019). We thank the participants at these meetings, and the anonymous reviewers, for their valuable comments.

## Funding

This work was supported by the Dutch Research Council (Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)) research talent grant for the project ‘Quality and inequality: The compensatory and multiplicative effects of school quality’ (NWO: 406-18-557, awarded to K.S., A.K., and I.M.) and a Veni grant for the project ‘Towards equal educational opportunities: The complex interaction between genes, families, and schools’ (NWO: 451-17-030, awarded to A.K.). We gratefully acknowledge research programme ‘Consortium on Individual Development’ which is funded through the Gravitation program of the Dutch Ministry of Education, Culture and Science and the NWO (NWO: 0240-001-003); ‘Decoding the gene–environment interplay of reading ability’ (NWO: 451-15-017); ‘Netherlands Twin Registry Repository: Researching the interplay between genome and environment’ (NWO: 480-15-001/674); ‘Twin-family study of individual differences in school achievement’ (NWO: 056-32-010); ‘Longitudinal data collection from teachers of Dutch twins and their siblings’ (NWO: 481-08-011); ‘KNAW Academy Professor Award’ (PAH/6635).

## Research Ethics

Informed consent was obtained from all participants. The study was approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre, Amsterdam, an Institutional Review Board certified by the U.S. Office of Human Research Protections (IRB number IRB00002991 under Federal-wide Assurance FWA00017598; IRB/institute codes, NTR 03-180).

## Data Availability

Netherlands Twin Register (NTR) is an ongoing longitudinal study. Data may be accessed upon reasonable request and after approval of the data access committee. For more information, see <https://ntr-data-request.psy.vu.nl/index.html>. Code for replication can be found here: <https://osf.io/mxzpe/>.

## References

- Akiba, M., LeTendre, G. K. and Scribner, J. P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. *Educational Researcher*, 36, 369–387.
- Alexander, K. L., Entwisle, D. R. and Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review*, 72, 167–180.
- Anderson, C. S. (1982). The search for school climate: a review of the research. *Review of Educational Research*, 52, 368–420.
- Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In Marcoulides, A. and Schumacker, R. E. (Eds.), *Advanced Structural Equation Modeling: Issues and Techniques*. Mahwah: Erlbaum, pp. 243–277.
- Atlay, C. et al. (2019). Instructional quality and achievement inequality: how effective is teaching in closing the social achievement gap? *Learning and Instruction*, 63, 1012111–1012110.
- Blatchford, P. and Russell, A. (2020). *Rethinking Class Size: The Complex Story of Impact on Teaching and Learning*. London: UCL Press.
- Blau, P. M. and Duncan, O. D. (1967). *The American Occupational Structure*. New York: Wiley.
- Boomsma, D. I. et al. (2002). Netherlands Twin Register: a focus on longitudinal research. *Twin Research*, 5, 401–406.
- Borghans, L., Golsteyn, B. H. H. and Zölitz, U. (2015). Parental preferences for primary school characteristics. *B.E. Journal of Economic Analysis and Policy*, 15, 85–117.
- Borman, G. D. and Kimball, S. M. (2005). Teacher quality and educational equality: do teachers with higher standards-based evaluation ratings close student achievement gaps? *Elementary School Journal*, 106, 3–20.
- Bourdieu, P., and Passeron, J.-C. (1977). *Reproduction in Education, Society and Culture*. London: Sage.
- Bradley, R. H. and Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, 53, 371–399.
- Bronfenbrenner, U. and Ceci, S. J. (1994). Nature-nurture reconceptualized in developmental perspective: a bioecological model. *Psychological Review*, 101, 568–586.
- Byrne, B. et al. (2010). “Teacher effects” in early literacy development: evidence from a study of twins. *Journal of Educational Psychology*, 102, 32–42.
- Centola, D. and Macy, M. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology*, 113, 702–734.
- Chiu, M. M. and Khoo, L. (2005). Effects of resources, inequality, and privilege bias on achievement: country, school, and student level analyses. *American Educational Research Journal*, 42, 575–603.
- Coenen, J. et al. (2018). Teacher characteristics and their effects on student test scores: a systematic review. *Journal of Economic Surveys*, 32, 848–877.
- Coleman, J. S. et al. (1966). *Equality of Educational Opportunity*. Washington: US Government Printing Office.
- Crosnoe, R. (2009). Low-income students and the socioeconomic composition of public high schools. *American Sociological Review*, 74, 709–730.

- de Graaf, N. D., de Graaf, P. M. and Kraaykamp, G. (2000). Parental cultural capital and educational attainment in the Netherlands: a refinement of the cultural capital perspective. *Sociology of Education*, 73, 92–111.
- de Zeeuw, E. L. et al. (2019). The moderating role of SES on genetic differences in educational achievement in the Netherlands. *NPJ Science of Learning*, 4, 1–8.
- DiPrete, T. A. and Eirich, G. M. (2006). Cumulative advantage as a mechanism for inequality: a review of theoretical and empirical developments. *Annual Review of Sociology*, 32, 271–297.
- Downey, D. B. and Condron, D. J. (2016). Fifty years since the Coleman report: rethinking the relationship between schools and inequality. *Sociology of Education*, 89, 207–220.
- Downey, D. B., von Hippel, P. T. and Broh, B. A. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review*, 69, 613–635.
- Dronkers, J. and Robert, P. (2008). Differences in scholastic achievement of public, private government-dependent, and private independent schools: a cross-national analysis. *Educational Policy*, 22, 541–577.
- Eifler, E. F., Starr, A. and Riemann, R. (2019). The genetic and environmental effects on school grades in late childhood and adolescence. *PLoS One*, 14, e02259461–e02259417.
- Engzell, P., Frey, A. and Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *PNAS*, 118, 1–7.
- Erola, J., and Kilpi-Jakonen, E. (Eds.). (2017). *Social Inequality Across the Generations. The Role of Compensation and Multiplication in Resource Accumulation*. Cheltenham: Edward Elgar Publishing.
- Franić, S., Dolan, C.V., Borsboom, D., and Boomsma, D.I. (2012). Structural equation modeling in genetics. In Hoyle, R. (Ed.), *Handbook of Structural Equation Modeling*. New York: Guilford Press, pp. 617–635.
- Grasby, K. L. et al. (2020). Estimating classroom-level influences on literacy and numeracy: a twin study. *Journal of Educational Psychology*, 112, 1154–1166.
- Greenwald, R., Hedges, L. V. and Laine, R. D. (1996). The effect of school resources on student achievement. *American Educational Research Association*, 66, 361–396.
- Gustafsson, J. E., Nilsen, T. and Hansen, K. Y. (2018). School characteristics moderating the relation between student socio-economic status and mathematics achievement in grade 8. Evidence from 50 countries in TIMSS 2011. *Studies in Educational Evaluation*, 57, 16–30.
- Hanselman, P. (2018). Do school learning opportunities compound or compensate for background inequalities? Evidence from the case of assignment to effective teachers. *Sociology of Education*, 91, 132–158.
- Hanushek, E. A. et al. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, 18, 527–544.
- Hanushek, E. A. and Rivkin, S. G. (2006). Teacher quality. In Hanushek, E. A. and Welch, F. (Eds.), *Handbook of the Economics of Education* (Vol. 2). Amsterdam: North-Holland, pp. 1051–1078.
- Hart, S. A., Little, C. and van Bergen, E. (2021). Nurture might be nature: cautionary tales and proposed solutions. *Npj Science of Learning*, 6, 1–12.
- ISLED. (2014). Overview of the International Standard Level of Education [ISLED]. Available from <http://www.harryganzeboom.nl/isled> [accessed 20 March 2021].
- Jones, L. and De Gioia, K. (2010). The same or separate? An exploration of teachers' perceptions of the classroom assignment of twins in prior to school and kindergarten to year two school settings. *Journal of Early Childhood Research*, 8, 239–253.
- Kahlenberg, R.D. (2001). The significance of the socioeconomic makeup of schools. In *All Together Now: The Case for Economic Integration of the Public Schools*. Washington: Brookings Institution, pp. 47–76.
- Kovas, Y. et al. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for Research in Child Development*, 72, 1–158.
- Kyriakides, L., Creemers, B. P. M. and Charalambous, E. (2019). Searching for differential teacher and school effectiveness in terms of student socioeconomic status and gender: implications for promoting equity. *School Effectiveness and School Improvement*, 30, 286–308.
- Lazear, E. P. (2002). Educational production. *The Quarterly Journal of Economics*, CXVII, 1–37.
- Lighthart, L. et al. (2019). The Netherlands Twin Register: longitudinal research based on twin and twin-family designs. *Twin Research and Human Genetics*, 22, 1–14.
- Ma, X. (2003). Sense of belonging to school: can schools make a difference? *The Journal of Educational Research*, 96, 340–349.
- Marks, G. N. (2021). Should value-added school effects models include student- and school-level covariates? Evidence from Australian population assessment data. *British Educational Research Journal*, 47, 181–204.
- Mönkediek, B. et al. (2020). Is there something special about twin families? A comparison of parenting styles in twin and non-twin families. *Social Science Research*, 90, 1024411–1024417.
- Morris, T. T. et al. (2018). Testing the validity of value-added measures of educational progress with genetic data. *British Educational Research Journal*, 44, 725–747.
- NVOM. (2019). Samen of apart? U zult het zelf moeten ontdekken. *Meerlingen Magazine*, Special Issue, 4–7.
- OECD. (2018). *Measuring Improvements in Learning Outcomes. Best Practices to Assess the Value-Added of Schools*. Paris: OECD.
- OECD. (2019). *PISA 2018 Results: Where All Students Can Succeed*. Paris: OECD.
- Parcel, T. L. and Dufur, M. J. (2001). Capital at home and at school: effects on student achievement. *Social Forces*, 79, 881–911.
- Plomin, R., DeFries, J.C., McClearn, G.E., and McGuffin, P. (2008). *Behavioral Genetics*. New York: Worth Publishers.
- Polderman, T. J. C. et al. (2010). No effect of classroom sharing on educational achievement in twins: a prospective, longitudinal cohort study. *Journal of Epidemiology and Community Health*, 64, 36–40.
- Preedy, P. (1999). Meeting the educational needs of preschool and primary aged twins and higher multiples. In Sandbank, A.C. (Ed.), *Twin and Triplet Psychology*. London: Routledge, pp. 70–99.

- Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research*, 5, 554–571.
- Rivkin, S. G., Hanushek, E. A. and Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73, 417–458.
- Robin, M. and Casati, I. (1994). Are twins different from singletons during early childhood? *Early Development and Parenting*, 3, 211–221.
- Rumberger, R. W. and Palardy, G. J. (2005). Does segregation still matter? The impact of student composition on academic achievement in high school. *Teachers College Record*, 107, 1999–2045.
- Scarr-Salapatek, S. (1971). Race, social class, and IQ. *Science*, 174, 1285–1295.
- Schröder, H. and Ganzeboom, H. B. G. (2014). Measuring and modelling level of education in European societies. *European Sociological Review*, 30, 119–136.
- Schwabe, I., Janss, L. and van den Berg, S. M. (2017). Can we validate the results of twin studies? A census-based study on the heritability of educational achievement. *Frontiers in Genetics*, 8, 1–8.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: a meta-analytic review of research. *Review of Educational Research*, 75, 417–453.
- Smith-Woolley, E. *et al.* (2018). Differences in exam performance between pupils attending selective and non-selective schools mirror the genetic differences between them. *NPJ Science of Learning*, 3, 1–7.
- Stanovich, K. E. (1986). Matthew effects in reading: some consequences of individual differences in the acquisition of literacy. *Research in Reading Quarterly*, 21, 360–407.
- Thapa, A. *et al.* (2013). A review of school climate research. *Review of Educational Research*, 83, 357–385.
- Turkheimer, E. *et al.* (2005). Analysis and interpretation of twin studies showing measures of the shared environment. *Child Development*, 76, 1217–1233.
- van Beijsterveldt, C. E. M. *et al.* (2013). The Young Netherlands Twin Register (YNTR): longitudinal twin and family studies in over 70,000 children. *Twin Research and Human Genetics*, 16, 252–267.
- van Leeuwen, M. *et al.* (2005). Effects of twin separation in primary school. *Twin Research and Human Genetics*, 8, 384–391.
- von Hippel, P. T., Workman, J. and Downey, D. B. (2018). Inequality in reading and math skills forms mainly before kindergarten: a replication, and partial correction, of “Are schools the great equalizer?” *Sociology of Education*, 91, 323–357.
- Wayne, A. J. and Youngs, P. (2003). Teacher characteristics and student achievement gains: a review. *Review of Educational Research*, 73, 89–122.
- Wentzel, K. R. and Ramani, G. B. (2016). *Handbook of Social Influences in School Contexts*. New York: Routledge.
- Kim Stienstra** is a PhD candidate at the Department of Sociology/ICS at Utrecht University. Her research interests include social stratification and inequality, and the interplay between genes and the environment. In her dissertation, she studies whether the school environment mitigates or exacerbates the influences of social origin, gender, and genes on children’s educational outcomes.
- Antonie Knigge** is an assistant professor in the Department of Sociology at Utrecht University and member of the ICS research school. His research interests are social stratification and mobility. He currently studies sources of sibling and twin (dis)similarity and the interplay between genes and the environment. He has published in *American Sociological Review*, *American Journal of Sociology*, and *Demography*.
- Ineke Maas** is a professor at the Departments of Sociology of Utrecht University and Vrije Universiteit Amsterdam. Her main research interests are international and historical comparisons of intergenerational mobility, career mobility, and marital mobility. In addition, she has published on the integration of immigrants, inequality of educational opportunities, and gender inequality.
- Eveline de Zeeuw** worked as an assistant professor at the Department of Biological Psychology (Vrije Universiteit Amsterdam) focusing on the association between psychopathology and educational achievement. During these years she was responsible for the data collection among teachers of twins at the Netherlands Twin Register.
- Dorret Boomsma** is a professor at the Department of Biological Psychology of Vrije Universiteit Amsterdam. She focuses on the causes of variation in human complex traits as a function of genotype and environment. To this end, she established the Netherlands Twin Register, which over the past 30 years recruited >200,000 twins and family members who take part in longitudinal phenotyping across the lifespan.