G Model
JSEE 612 No. of Pages 15

ARTICLE IN PRESS

Studies in Educational Evaluation xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/stueduc

# School characteristics moderating the relation between student socio-economic status and mathematics achievement in grade 8. Evidence from 50 countries in TIMSS 2011

Jan-Eric Gustafsson[a,b,*], Trude Nilsen[c], Kajsa Yang Hansen[a]

[a] University of Gothenburg, Department of Education and Special Education, P.O. Box 300, S-40530 Gothenburg, Sweden
[b] University of Oslo, CEMO, Postboks 1161, Blindern, Oslo 0318, Norway
[c] University of Oslo, Department of Teacher Education and School Research, P.O. Box 1099, Blindern, 0317 Oslo, Norway

ARTICLE INFO

*Article history:*
Received 14 December 2015
Accepted 21 September 2016
Available online xxx

*Keywords:*
Socioeconomic status
Equal education
Mathematics achievement
Educational quality
School climate
School segregation
TIMSS
Random slopes modeling

ABSTRACT

The main aim of the study was to identify school characteristics that can reduce the relation between socio-economic status (SES) and achievement, so that equity of educational outcomes can be improved. Data from 50 countries participating in the Trends in International Mathematics and Science Study (TIMSS) conducted in 2011, focusing on Grade 8 mathematics, was analysed. Two-level random slopes models fitted at school- and student-levels were used to investigate the influence of quality and quantity of instruction, school climate, and school SES on the within-school regression slope for achievement on SES. The results showed school SES to be the strongest determinant of slope differences across schools and educational systems. Whether school SES relates negatively or positively to the within-school regression of achievement on student SES is an indicator of whether the educational system is compensatory or anti-compensatory with respect to student SES.



## 1. Introduction

Socio-economic status (SES) refers to an individual's or a family's position in a hierarchy according to access to wealth, power, and social status. The correlation between students' academic achievement and family SES is in most countries around 0.20–0.40 at the individual level (Sirin, 2005) and with data aggregated to the class or school levels it is considerably higher. In order to increase educational equity, the strength of these relations needs to be reduced. However, little is known about which school factors influence the relationship between SES and educational achievement. One reason for this may be that little attention has been devoted to investigating the mechanism through which SES is related to educational achievement. Instead researchers have taken advantage of SES to control for selection bias in investigations of effects of school factors and instructional variables, thus focusing on main effects of SES. However, if the aim is to influence the strength of the relationship between SES and educational achievement, school characteristics that reduce the relation between SES and achievement need to be identified.

One challenge when investigating effects of school factors in observational studies is that the amount of variation in the investigated factors often is restricted within any particular country. However, taking advantage of international comparative large-scale data may increase the possibility of identifying factors influencing the strength of the relationship between SES and achievement. The current study is based on data from 50 countries participating in the Trends in International Mathematics and Science Study (TIMSS) conducted in 2011, focussing on outcomes in the area of mathematics. We use these data to investigate the influence of quality and quantity of instruction, school climate, and school SES on the relation between SES and achievement.

### 1.1. Conceptualization of SES

A large body of empirical evidence has established student family SES as being one of the most powerful predictors of school outcomes (*e.g.*, Sirin, 2005; White, 1982). However, there is little consensus on precisely what SES represents (*e.g.*, Liberatos, Link, & Kelsey, 1987; McLoyd, 1998), and there is great variation in the relationship between SES and educational achievement across different studies. Sirin (2005) concluded that the variation in the

* Corresponding author at: University of Gothenburg, Department of Education and Special Education, P.O. Box 300, S-40530 Gothenburg, Sweden.
*E-mail addresses:* Jan-Eric.Gustafsson@ped.gu.se (J.-E. Gustafsson), trude.nilsen@ils.uio.no (T. Nilsen), kajsa.yang-hansen@ped.gu.se (K.Y. Hansen).

http://dx.doi.org/10.1016/j.stueduc.2016.09.004
0191-491X/© 2016 Published by Elsevier Ltd.

strength of the SES effect may partially be accounted for by how SES is being measured. Family SES is typically measured by parental education level, occupation and income but other indicators, such as eligibility for free lunch, or material possessions, number of siblings, family structure and ethnicity are also used. Most often, SES is measured as a composite of different indicators, reflecting the view that SES may be regarded as a combination of different types of capital or resources that influence children's development (Coleman, 1988, Bourdieu, 1986). It has been argued that measuring SES as a unidimensional construct may neglect some of the important aspects of SES (e.g., Yang & Gustafsson, 2004; Yang, 2003). Using Programme for International Student Assessment (PISA) 2000 data, Marks, Cresswell and Ainley (2006) found that family cultural resources played a more important role than family material resources, and that cultural resources, such as number of books in the home, explained a substantial amount of socioeconomic inequality in academic achievement in most of the countries. To explain differences in educational achievement, the best approach to measure student SES thus is to use a relatively broad measure of home educational resources that captures aspects such as number of books at home, parental level of education and amount of study support at home. Such scales have been constructed, for example, for TIMSS (Martin, Mullis, Foy, & Arora, 2012) and for PISA (OECD, 2012).

### *1.2. The relations between SES and educational achievement*

While a pervasive influence of SES on achievement has been demonstrated in numerous studies, explanatory models which account for the mechanisms through which the influence occurs are not so well developed. Within educational research, there are basically two ways in which such factors can cause SES to be more or less strongly related to educational achievement. First, educational factors may have differential effects on low- and high-SES students. For example, if instructional quality has a stronger positive effect on the achievement of low-SES than high-SES students, there is a differential or interactive effect. Another way to phrase this is to say that instructional quality moderates the relation between SES and achievement. The second way that educational factors may influence the observed relationship between SES and achievement is through a correlation between SES and educational factors. For example, if low-SES students tend to be provided with instruction of lower quality than high-SES students, this will cause their level of achievement to be lower. The effect will be a function both of the amount of difference in level of quality of instruction between the SES-groups, and of the extent to which quality of instruction is related to achievement. These are additive effects of SES and instructional quality, and we conceive of instructional quality as a factor which mediates the effect of SES on achievement.

Moderating and mediating mechanisms may operate simultaneously, and it is easy to imagine scenarios in which the two mechanisms either reinforce or counteract one another. This may be one of the reasons why it is a difficult task to sort out which factors influence the relation between SES and achievement. Below we first review the effect of collective SES on achievement, and then we review other factors influencing the relationship between family SES and achievement.

### *1.3. Equity and the effect of collective SES*

Given that classrooms, schools and neighbourhoods differ with respect to the SES of their members, we can think of a collective SES, defined within a multi-level framework as the mean level of SES of the members of the group (Yang & Gustafsson, 2004). School-SES or collective SES may exert both additive and interactive effects on educational achievement. Rutter, Maughan, Mortimore and Ouston (1979) concluded that there were effects of schools as social institutions even after the student SES was controlled for. Such school differences may, or may not, be due to collective SES, but given the strong relations between school SES and school achievement typically found (Sirin, 2005), it is necessary to take school SES into account.

Collective SES shapes the overall learning environment through its association with social mechanisms and factors which may influence educational outcomes at different levels. These mechanisms include social stratification, peer effects, contextual effects, educational choice and self-selection, as well as institutional differentiation (e.g., Coleman, 1988 Thrupp, 1999; Thrupp, Lauder & Robinson, 2002; Van de Werfhorst, & Mijs, 2010). Numerous studies have concluded that students who attend low-SES schools perform worse than students who attend high-SES schools, even after controlling for students' family background and their ability upon entry to school (e.g., Liu, Van Damme, Gielen, & Van Den Noortgate, 2015; Palardy, 2013; Schmidt, Burroughs, Zoido, & Houang, 2015; Van Ewijk & Sleegers, 2010). Thus, there are reasons to assume that the disparity in educational outcomes of different schools is partially determined by differences in the social and institutional factors that are associated with school SES, over and above effect of individual SES.

Van Ewijk and Sleegers (2010) found in a meta-analysis a substantial variation in estimates of effects of school SES. They argued that the lack of consensus may partly be due to methodological issues related to operationalization of SES, and partly to lack of control for omitted variable bias.

Previous research (e.g., OECD, 2013), has shown that equitable educational systems tend to achieve better results than non-equitable educational systems. Different indicators of equity have been used, such as dispersion of student achievement, amount of school differences in achievement, the within-school regression of student achievement on student SES, and the between-school regression of school-level achievement on school-SES. It is important to determine the characteristics of different measures of equity, and how they relate to country level achievement.

### *1.4. School factors influencing the relation between student SES and achievement*

Up until the mid-1990s the prevailing view among many groups of researchers, and particularly among economists, was that resources matter little for educational outcomes (Burtless, 1996; Hanushek, 1989). However during the last couple of decades new methods for synthesizing results from different studies and an increased number of high-quality studies have changed this negative view. Using meta-analytic techniques Greenwald, Hedges, and Laine (1996) concluded that there is quite a strong relationship between school resources and educational results. Several studies also found that the effects of resources, such as class-size, were stronger for low SES students than high SES students (e.g., Finn & Achilles, 1999; Krueger, 2003; Nye, Konstantopoulos, & Hedges, 2004; Wenglinsky, 1998; see also more recent reviews on class-size effects by Ehrenberg, Brewer, Gamoran, & Willms, 2001 and by Ecalle, Magnan & Gilbert, 2006, which present more complex patterns of results). One interpretation of the interactive effect was that in schools lacking adequate resources to compensate low SES students for their less adequate preparation, the outcomes will to a larger extent be based on the students' family background (e.g. Wenglinsky, 1998). A recent reanalysis of the Coleman data by Borman and Dowling (2010) investigated the effects of school-level SES and school resources. They showed that within-school variation of achievement is explained by ability tracking and

Please cite this article in press as: J.-E. Gustafsson, et al., 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 (2016), http://dx.doi.org/10.1016/j.stueduc.2016.09.004

teacher effects. These results imply that schools and classrooms matter in explaining student achievement and the effect of SES (; see also e.g., Perry & McConney, 2010).

School climate creates the premises for instruction and learning and has been found to influence the relation between SES and achievement (e.g., Johnson & Stevens, 2006; Kyriakides, Creemers, Antoniou & Demetriou, 2010; Papanastasiou, 2008; Uline & Tschannen-Moran, 2008). In a review of research on school climate, Wang and Degol (2015) observed that school climate is defined differently across studies, but that certain aspects are more important than others. One key aspect of school climate is the priority and ambition for learning and success (Hoy, Tarter & Hoy, 2006). Research on the construct School Emphasis on Academic Success (SEAS), which reflects such priority (Hoy et al., 2006), has been shown to be related to students' learning in a number of countries (Martin, Foy, Mullis, & O'Dwyer, 2013; Nilsen & Gustafsson, 2014). Lee and Smith (1999) found that the SEAS effect was more powerful in low SES schools than in high SES ones).

A safe and orderly climate is another key aspect of school climate, safety referring to the degree of physical and emotional security provided by the school, and an orderly climate to disciplinary practices (Goldstein, Young, & Boyd, 2008; Gregory, Cornell, & Fan, 2012; Wang & Degol, 2015). This aspect of school climate is positively related to student outcomes, and also to school SES (e.g. Bryk & Schneider, 2002; Hoy et al., 2006; Thapa, Cohen, Guffey, & Higgins-D'Alessandro, 2013). These studies thus suggest that school SES is strongly related to a healthy and beneficial school climate. One explanation could be that high SES students and their parents are more likely to both contribute to and demand positive school climates (e.g. Brantlinger, 2003). These results indicate that there are close connections between school SES and school climate which creates challenges in disentangling the unique effects of these two factors.

Few studies have directly investigated the mechanisms through which school factors influence socioeconomic inequality of educational achievement. However, in a recent study, Liu, Van Damme, Gielen, and Van Den Noortgate (2015) examined whether school processes mediate the effect of school SES on mathematics literacy. Using PISA data they found that school SES effects were partially mediated by school climate factors. Also Rumberger and Palardy (2005) found in a longitudinal study that school composition effects were partly mediated by school process factors, such as school climate. Other studies applying different modelling approaches and data sources have reported similar findings (e.g., Opdenakker & Damme, 2007).

Instructional quality, comprising aspects such as cognitive activation, supportive climate, clarity of instruction, and classroom management, has been shown to have a positive influence on student achievement (Baumert et al., 2010; Hiebert & Grouws, 2007; Klieme, Pauli & Reusser, 2009; Pianta & Hamre, 2009). Rjosk et al. (2014) investigated the mediation of SES via instructional quality in a two-year German longitudinal study. The results indicated that focus on cognitive activation in the form of challenging language instruction partly mediated the classroom SES effect on achievement. The mechanism behind the mediation effect seemed to be that teachers focused less on challenging language instruction in low SES classrooms.

Willms (2010) used data from PISA 2006 to examine the relationships among school SES, different aspects of the school and classroom contexts, including instructional quality, and students' science literacy skills. School SES effects were mediated by the quality of instruction and time allocated to science instruction, students in low SES schools having considerably less time spent on science instruction than those in high SES schools.

Previous research has shown instructional time to be a powerful predictor of school achievement (Good, Wiley, & Florez, 2009, p. 806; Scheerens, 2014). Like school climate and instructional quality, also quantity of schooling tends to be related to SES, low SES students being allocated less time on task (Burger, 2016; Oakes, 1985). Schmidt et al. (2015) investigated how opportunity to learn may mediate the effect of SES on mathematics literacy using data from PISA 2012. They found that roughly a third of the SES relationship to mathematics literacy was due to differences in opportunity to learn and emphasized that this indicates that SES-related inequalities in school outcomes may not only be a product of student characteristics and home background but also of schooling itself. However, they also found that some educational systems managed to reduce SES inequalities, which suggests that it is possible to organize schooling in such a way that quality and quantity of schooling can partially compensate for less favourable background conditions.

### *1.5. Conclusions and research questions*

The research reviewed above supports the hypothesis that there are factors which moderate the relation between SES and achievement. Several studies indicate that instructional quality and quantity, and aspects of school climate lower the relation between SES and achievement, because the positive effect is stronger for low-SES students than for high-SES students. The hypothesis that SES is involved in mediating relations is also supported and it has typically been concluded that low-SES students are provided with less quantity of instruction, instruction of lower quality and poorer school climate than are high-SES students. These results suggest that improving school climate and the quality and quantity of instruction for low-SES students could have substantial effects on the equity of educational outcomes.

The studies also indicate that there are very high correlations between school SES on the one hand, and school climate and other school characteristics on the other hand. This correlation makes it challenging to separate the effects of school SES, school climate and instructional quality and to identify the mechanisms through which they exert their effects. It could be that school SES generates a positive school climate, in which case the school SES effect should be seen as being mediated via school climate. But other views are, of course, also possible, such as the traditional one in which school SES is controlled for statistically. When implemented in analytic models, these views lead to very different results, which is due to the fact that the models applied often are based on naïve or incorrect assumptions, such as when analysis of covariance is used to control for pre-existing differences (Miller & Chapman, 2001) or when cross-sectional mediation modelling is used to investigate longitudinal mediational processes (Maxwell, Cole & Mitchell, 2011).

Increasingly, the large-scale international studies of educational achievement, and particularly so PISA, have been taken advantage of in the research on educational equity. However, while this is an excellent source of comparative data, our review reveals that most studies have been conducted on western, developed, countries like the Organisation for Economic Co-operation and Development (OECD) countries, while few studies have included developing countries. Given that it should be informative to investigate educational equity not only within countries but also across countries this is a limitation. A related limitation is that only few studies have considered both equity and efficiency of education. While there are indications that equity of education is associated with a higher level of achievement (OECD, 2013; Van de Werfhorst & Mijs, 2010) there is little research on the generality of this relation, or on why there is such a relation.

Please cite this article in press as: J.-E. Gustafsson, et al., 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 (2016), http://dx.doi.org/10.1016/j.stueduc.2016.09.004

As has been demonstrated by among others Schmidt et al. (2015), effects of SES on achievement are due both to processes within schools and to differences between schools. However, most of the research has focused on between-school differences. Following the lead of Bryk and Raudenbush (1988) the present study has been designed to focus on school differences in within-school relations between student SES and achievement and compare these across educational systems. We also investigate if school characteristics (quality and quantity of instruction, school climate and school SES) can account for the variation in the within-school relations. The modelling is repeated for all 50 participating educational systems in the grade 8 TIMSS 2011 study (Mullis, Martin, Foy, & Arora, 2012) and the parameter estimates from the 50 models are assembled to allow analyses at the educational system level.

The following research questions are investigated:

1. To what extent do within-school regressions of mathematics achievement on student SES vary across schools and to what extent do mean slopes vary across educational systems?
2. To what extent can slope differences be accounted for by school characteristics reflecting quality and quantity of instruction, school climate, and school SES?
3. How do differences in equity relate to level and dispersion of mathematics achievement across educational systems?

## 2. Method

In the study we reanalyse TIMSS 2011 data by applying two-level random slopes modelling at school and student levels. The data and procedures used in the analyses are described below.

### 2.1. The database

TIMSS is an international large scale survey of students' mathematics and science achievement in grade 4 and 8 conducted by the International Association for the Evaluation of Educational Achievement (IEA). The present study includes all students ($N = 282\ 737$) from the 50 countries that participated in grade 8 in TIMSS 2011 (Mullis et al., 2012).

TIMSS includes representative samples of students nested within classes that are nested within schools. Students, teachers and principals answer questionnaires, providing background and contextual information. Most of the independent variables used in the analyses were created from this information.

### 2.2. Variables

#### 2.2.1. Mathematics achievement

The dependent variable mathematics achievement was measured by items in the domains of algebra, geometry, number, and data and chance. Altogether there were around 200 items, but TIMSS uses a matrix-sampling design according to which each student was administered only a subset of the test items. All achievement scores were expressed on a common scale in the form of so called 'plausible values' which are multiple imputed scores, taking advantage of all available responses to both test items and background variables (see, e.g., von Davier, Gonzalez & Mislevy, 2009). There were five plausible values in mathematics but the analysis was restricted to the first plausible value.

#### 2.2.2. SES

A variable derived from several items in the student questionnaire (students' ratings of the number of books at home, their parents' highest education and home study supports such as students having their own room and internet connection) was available in the TIMSS data base to represent students' SES (the Home Educational Resources scale). TIMSS used a partial credit model to estimate the scale (Martin et al., 2012). This variable was used as a single observed manifest variable at both student (SES) and school levels (School-SES).

#### 2.2.3. Instructional quality

Students' ratings were used to measure Instructional quality (InQua). Three statements pertaining to the question: "How much do you agree with these statements about your mathematics lessons?" were rated on a four point Likert scale ranging from "disagree a lot" to "agree a lot". The statements were "I know what my teacher expects me to do", "My teacher is easy to understand", and "I am interested in what my teacher says". These three variables were used as indicators in a two-level model of the student-level latent variable InQuaW to represent within-school differences between students in the ratings of quality of instruction, and the school-level latent variable InQua, to represent shared perceptions of InQua in each school (Marsh et al., 2012; Scherer & Gustafsson, 2015).

#### 2.2.4. Instructional quantity

We included one aspect of opportunity to learn, namely time allocated for instruction. This is an index available in the TIMSS data base, which is based on the principals' responses to the following questions: "How many days per year is your school open for instruction?" and "What is the total instructional time, excluding breaks, in a typical day?" Open responses were asked for, with a range between 486 and 2327 h per year, and a mean of 1041 (SD 222) hours. This variable was used as an observed school-level variable.

#### 2.2.5. School climate

As described above two important aspects of school climate are School Emphasis on Academic Success (SEAS) and a Safe and Orderly climate. The SEAS construct was rated by school leaders' answers to the question: "How would you characterize each of the following within your school?" and three statements concerning teachers' understanding of the curriculum, their success with the curriculum, and teachers' expectations for students' success. Responses were given on a five point Likert scale. These three variables were used to identify the school-level latent variable SEAS.

A Safe and Orderly climate was measured by school leaders' answer to the following question: "To what degree is each of the following a problem among the students in your school?" Ten statements were rated on a four-point Likert scale. The statements concerned: arriving late at school, absenteeism, classroom disturbance, cheating, profanity, vandalism, theft, intimidation or verbal abuse among students, physical injury to other students, and intimidation or verbal abuse of teachers or staff. These ten variables were used to identify the school-level latent variable Order.

#### 2.2.6. Human development index

The HDI is a general measure of country level human development (UNDP, 2014). The index is a composite measuring three basic dimensions: life expectancy at birth; schooling, measured both as mean years of schooling received by persons aged 25 and older and expected years of schooling for a child at school entrance; and gross national income per capita. HDI values for several years are available (UNDP, 2014), and we used the HDI values for 2012. The variable was not available in the TIMSS data, so it was added to the dataset and was used as an observed control variable in country level analyses.

Please cite this article in press as: J.-E. Gustafsson, et al., 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 (2016), http://dx.doi.org/10.1016/j.stueduc.2016.09.004

### 2.3. Statistical analyses

The International Database (IDB) analyzer (2015) was used to merge the student- and school-level data, and Mplus 7.3 (Muthén & Muthén, 1998–2014) was relied upon for the statistical analyses. In order to handle missing data, the full information maximum likelihood procedure implemented in Mplus was applied. In all analyses, we used the robust maximum likelihood estimator (MLR).

To address our research questions, we specified two-level random slopes structural equation models, using school as the between-level. The model assumed the within-school estimates of the slope and intercept for the regression of mathematics achievement on SES to be random coefficients, so the model estimated the mean and the variance of the slope and the intercept. The idea thus is that within some schools the relationship between individual SES and their math achievement (i.e., the slope parameter) is stronger than in other schools, and that we can capture these differences with the estimated variance of the slopes.

These estimates, along with the estimates of the mean of the within-school slopes from each country, are educational system level characteristics that can be related to other system level characteristics. The two-level random slopes model can also be extended by adding school characteristics in a school-level regression model to explain variation in the within-school slopes and intercepts. The regression coefficients from such analyses also are characteristics at the educational system level.

Given that we are interested in differences among educational systems we could, in principle, have specified a three-level model, with student, school, and educational system as the three levels. However, this option was not supported by Mplus 7.3 (Muthén & Muthén, 1998–2014), and it would have resulted in an extremely large and complex model, We therefore first computed estimates for the two-level random slopes null model for each of the 50 countries, regressing mathematics achievement on student SES within schools without including any school-level predictors. We refer to the latent variable representing the within-school slopes as Slope. The analyses also yielded estimates of the mean and variance of Slope for each educational system, which we refer to as Slope_mean and Slope_var, respectively.

In the next step, analyses were done, aiming to investigate to what extent the school-level variability in the within-school achievement—SES slope can be explained by school characteristics (i.e., InQua, Hours, SEAS, Order and School-SES). This is thus an investigation of cross-level interaction, or if school characteristics moderate the within-school relationship between student SES and achievement. A separate analysis was conducted for each school characteristic (or “moderator”) and each analysis was run for all 50 educational systems in a multiple-group model, with equality constraints of factor loadings across groups for the latent variables, if any. Hence, five different models were estimated, yielding estimates for the 50 educational systems. We refer to the estimated coefficients of the moderators as Slope_InQua, Slope_Hours, Slope_SEAS, Slope_Order, and Slope_School-SES. In a few cases, however, the model failed to converge for a single educational system, and in these cases no estimate was obtained.

The models also included the regression of the school-level intercept of achievement on the school characteristic. This part of the model thus estimates the between-school regression of achievement on the school characteristics. Among these we focus on the relation between the school-level intercept and school-SES, which is used as a measure of equity at the country level. This estimated coefficient will be referred to as Int_School-SES.

The path-diagram for an example of the two-level models estimated is shown in Fig. 1.

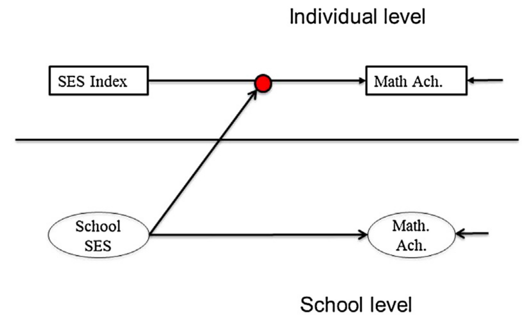


**Fig. 1.** Path diagram for one of the two-level models, with a random slope within schools for the regression of mathematics achievement on SES, and the random slope regressed on school SES.

The main focus of the analyses thus was to investigate to what extent the variability in Slope within systems can be moderated by school characteristics. A negative regression coefficient for the moderator InQua, for example, would imply that schools with a higher student rated quality of instruction have a lower impact of student SES on mathematics achievement than schools with a lower level of quality. Conversely, if Slope_InQua is positive, a higher student rated quality of instruction is associated with a stronger correlation between student SES and math achievement and thus with a lower level of equity within schools.

Research question 3 asks how level and dispersion of mathematics achievement across educational systems relate to differences in equity. In this research, different indicators of equity have been relied upon. One is the standard deviation of student achievement, and another is the proportion of between school variance in achievement in relation to the total variance (i.e., the Intraclass Correlation Coefficient, ICC). A third measure is the between-school regression of achievement on school-SES which OECD (2012) refers to as the School SES Gradient and which we refer to as Int_School-SES (see above). We also expect the moderator Slope_School-SES to be an indicator of equity. We thus have four measures of equity at the educational system level (i.e. the ICC, Math_SD, Int_School-SES and Slope_School-SES) and it is of great interest to compare these measures of equity when it comes to prediction of mean achievement. It also is of interest to investigate predictors of Math_SD.

### 2.4. Descriptive statistics and estimates of model parameters

Appendix A presents descriptive statistics for Mathematics achievement and SES at student and school levels, along with information about numbers of students and schools, ICC for school differences in achievement, and HDI. The analyses yielded a large number of estimates, which are presented at the educational system level in Appendix B. The data presented include estimates of Slope_mean, Slope_var and coefficients for the regression of Slope on the five school characteristics, along with Int_School-SES. These data form the basis for the country-level analyses, which have been done with simple statistical techniques, such as bivariate correlations and scatter plots, along with regression analyses.

However, before the results from these analyses are presented, we will make a few comments concerning the estimates presented in Appendix B. The first two columns are unstandardized estimates of mean and variance of the Slope for each country. It may be observed that the Slope means varied considerably across the educational systems, there being a clear tendency for the developing countries to have lower within-school regression coefficients of math achievement on student SES.

The Slope variances tended to be small and in most cases they were non-significant. However, for about a dozen educational systems, many of which were in developing countries, they were significant at the 0.05 level. In these systems, estimates typically were higher than 0.009. Nevertheless, even for educational systems with small estimated variance, it frequently was the case that Slope was significantly predicted by one or more of the school characteristics. This was observed for some 20 educational systems, suggesting that the variance estimates obtained in the null models without any predictor variables were estimated with low precision, and that adding school-level predictor variables increased precision.

The estimates of the coefficients for the regression of Slope on school characteristics are presented in the form of t-values (i.e., the parameter estimates divided by its standard error). One reason for this was that in some instances both parameter estimates and standard errors were large, creating outliers in the data. However, the t-values generally were within reasonable bounds. These estimates also have the advantage of being easy to interpret with respect to statistical significance, a rough rule of thumb being that absolute vales larger than 2 may be regarded as significant.

Some of the values in Appendix A and B were missing. This was the case for the HDI, which was not available for two countries. In a few cases the iterative algorithm for estimating the random-slopes model also failed to converge, and no estimates were obtained.

## 3. Results

Table 1 presents correlations among the variables presented in Appendix B, and these correlations will form our starting point in answering the three research questions.

### 3.1. Estimates of within-School regressions

Our first research question asked to what extent within-school regressions of mathematics achievement on student SES differ across schools for each of the 50 educational systems, as reflected in the variance and the mean of the Slope. As has already been observed the estimate of the variance of the within-school regression Slope variable was significant for about a dozen educational systems, many of which were in developing countries. Nevertheless, a significant amount of Slope variance was predicted for an additional 20 educational systems, which indicates that heterogeneity of regressions can more easily be detected when school-level predictors are included in the model.

There was a strong negative correlation (−0.47) between the Slope variance and the HDI, and there was an equally strong positive correlation between the Slope mean and the HDI. Thus, while developing countries had a larger Slope variance, the Slope mean was lower. This pattern may reflect differences in school organization between the high-HDI countries and the low-HDI countries, there being less organizational differentiation and more comprehensive schooling in high-HDI countries.

### 3.2. Correlations between estimates from the slope regressions

Our second research question asked to what extent school characteristics reflecting quality and quantity of instruction, school climate, and school SES moderate the within-school relation between student SES and achievement. The regression coefficients representing the amount of influence of each of the five school characteristics on the relation between individual SES and achievement (Slope) had a pattern of correlations that grouped them into two categories. The regression coefficients of Slope on the two moderators instructional quality (Slope_InQua) and instructional quantity (Slope_Hours) were positively correlated

**Table 1**
Correlations among regression parameters and other characteristics of the educational systems.

| | Achievement | Dispersion of achievement | | | The slope | | Moderators | | | | | Human Development Index |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Math_mean | Math_SD | ICC | Int_School-SES | Slope mean | Slope var | Slope_InQua | Slope_Hours | Slope_SEAS | Slope_Order | Slope_School_SES | HDI |
| Math_mean | 1 | | | | | | | | | | | |
| Math_SD | −0.24 | 1 | | | | | | | | | | |
| ICC | −0.19 | 0.21 | 1 | | | | | | | | | |
| Int_School-SES | −0.07 | 0.14 | 0.63$^{**}$ | 1 | | | | | | | | |
| Slope mean | 0.49$^{**}$ | 0.23 | −0.56$^{**}$ | −0.21 | 1 | | | | | | | |
| Slope var | −0.52$^{**}$ | 0.28 | 0.15 | 0.01 | −0.30$^{*}$ | 1 | | | | | | |
| Slope_InQua | −0.24 | −0.08 | −0.12 | −0.17 | 0.01 | 0.11 | 1 | | | | | |
| Slope_Hours | −0.15 | −0.15 | 0.012 | −0.13 | −0.21 | 0.01 | 0.37$^{**}$ | 1 | | | | |
| Slope_SEAS | −0.49$^{**}$ | 0.16 | 0.09 | 0.19 | −0.32$^{*}$ | 0.36$^{*}$ | 0.19 | 0.21 | 1 | | | |
| Slope_Order | −0.34$^{*}$ | 0.35$^{*}$ | 0.08 | 0.17 | −0.11 | 0.12 | 0.03 | −0.06 | 0.37$^{**}$ | 1 | | |
| Slope_school SES | −0.64$^{**}$ | −0.01 | 0.16 | −23 | −0.48$^{**}$ | 0.52$^{**}$ | 0.05 | 0.03 | 0.57$^{**}$ | 0.41$^{**}$ | 1 | |
| HDI | 0.79$^{**}$ | −0.37$^{*}$ | −0.13 | 0.04 | 0.47$^{**}$ | −0.47$^{**}$ | −0.20 | −0.25 | −0.40$^{**}$ | −0.41$^{**}$ | −0.50$^{**}$ | 1 |

Note. Math mean = Mean of TIMSS 2011 mathematics achievement (first plausible value); Math_SD = standard deviation of TIMSS 2011 mathematics achievement (first plausible value); ICC = Intraclass correlation (proportion of between school variance in achievement in relation to the total variance); Int_School-SES = coefficient for the regression of school mean achievement on School-SES; Slope mean = Mean of within-school regression slopes for student math score on SES; Slope var = Variance of within-school regression slopes for student math score on SES; Slope_InQua = t-value for the regression of the within-school Slope on school-level latent variable representing instructional quality; Slope_Hours = t-value for the regression of the within-school Slope on school-level yearly hours of instruction; Slope_SEAS = t-value for the regression of the within-school Slope on school-level latent variable representing School Emphasis on Academic Success; Slope _Order = t-value for the regression of the within-school Slope on school-level latent variable representing Safe and Orderly school; Slope_School SES = t-value for the regression of the within-school Slope on school SES; HDI = Human Development Index 2012, as defined by (UNHDP, 2014).

$^{*}$ Correlation is significant at the 0.05 level (2-tailed).

$^{**}$ Correlation is significant at the 0.01 level (2-tailed).

(0.37), but both had low correlations with the other coefficients. The regression coefficient of Slope on the moderator School-SES (Slope_School-SES) was positively correlated with that of the moderator SEAS (Slope_SEAS) (0.57) and the moderator reflecting a safe and orderly climate (Slope_Order) (0.41). The latter two also were positively correlated (0.37). Hence two groups of moderators emerged, one group including indicators of quality and quantity of instruction, and another group including indicators of school organization and school climate.

While this pattern of correlations is interesting, it is not informative about the characteristics of individual educational systems. Fig. 2 presents a scatter plot between the t-values of the regression coefficients of Slope on the t-values of the moderators instructional quality (Slope_InQua) and School-SES (Slope_School-SES) for the 50 educational systems. These two variables were essentially uncorrelated, but the scatter plot provides information both about the combinations of values for the two variables for each educational system and about the distributions of values for each of the two variables.

The School-SES variable had by far the largest number of significant estimates according to the rough rule of thumb that absolute values larger than 2 are significant (see also Appendix B). For 10 educational systems, School-SES had a significant and positive influence on Slope and for equally many educational systems it was significant and negative. The group of educational systems with positive regression coefficients was primarily composed of those with large slope variance (South Africa, Tunisia, Morocco, Botswana, Turkey, Iran, Indonesia, Thailand, Ghana and Honduras). The group with negative coefficients included East Asian educational systems (Chinese Taipei, Japan, and Singapore), Eastern European/former Soviet Union educational systems (Armenia, Georgia, Hungary, Lithuania, Russian Federation and Ukraine), along with Canada (Quebec). A positive coefficient implies that a high level of school-SES is related to a steeper within-school SES-achievement relation, which indicates that the system is anti-compensatory with respect to student SES. A negative coefficient indicates, in contrast, that the educational system is compensatory with respect to student SES.

Three educational systems had significant negative values for the regression of Slope on instructional quality (Slope_InQua) (Singapore, Hong Kong and Chinese Taipei). A negative coefficient implies that a higher level of instructional quality is related to a less steep within-school slope of achievement on student SES, i.e., to a higher degree of equity. Hungary and Armenia, in contrast, had large positive coefficients for Slope_InQua, so in these educational systems, a higher level of instructional quality was related to a steeper within-school slope of achievement on student SES, i.e. to a lower degree of equity.

Fig. 3 presents a scatter plot between the regression coefficients for the moderators SEAS (Slope _SEAS) and School-SES (Slope_-School-SES). These two variables were positively correlated (0.57), and the scatter plot provides information about the combination of values for the two variables for each educational system. The results presented for Slope_School-SES are of course identical to those presented above.

Four educational systems had negative values for Slope_SEAS (Australia, Chinese Taipei, Lithuania and Canada, Quebec). A negative coefficient implies that a school climate with larger emphasis on academic success is related to a less steep within-school slope of achievement on student SES, i.e., to a higher degree of equity. South Africa and Israel, in contrast, had large positive coefficients for Slope_SEAS, so in these educational systems a larger emphasis on academic success was related to a steeper within-school slope of achievement on student SES, indicating a lower degree of equity.

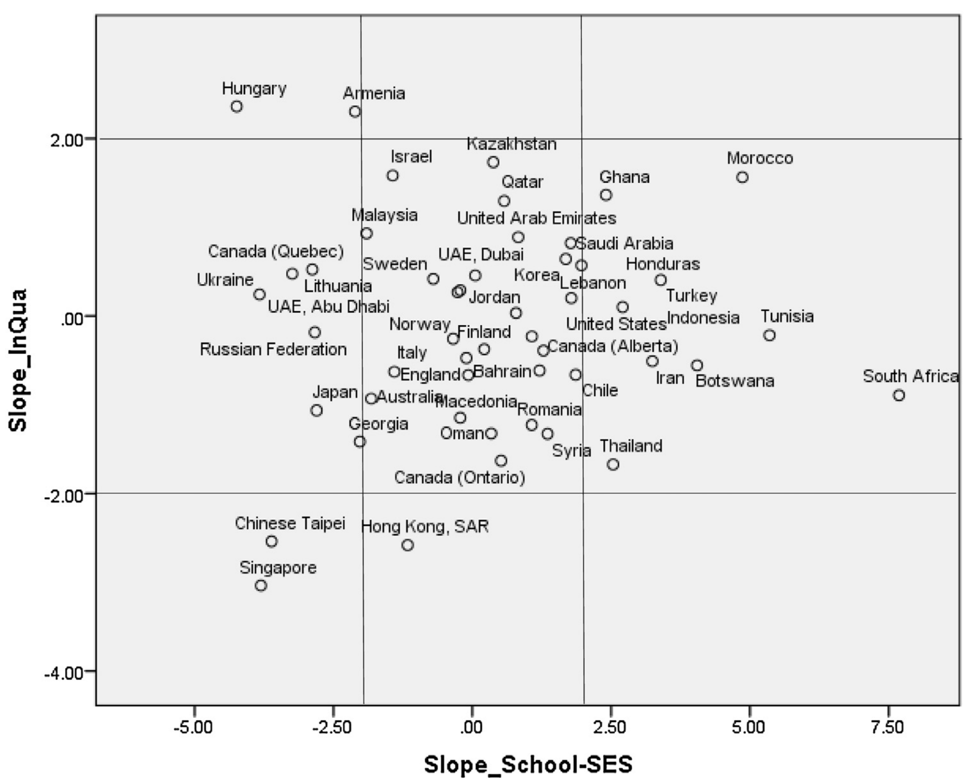


**Fig. 2.** Plot of Slope_InQua with Slope_School-SES.
*Note*: Slope_InQua = t-value for the regression of the within-school random slope on school-level latent variable representing instructional quality; Slope_School-SES = t-value for the regression of the within-school random slope on school SES.

Please cite this article in press as: J.-E. Gustafsson, et al., 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 (2016), http://dx.doi.org/10.1016/j.stueduc.2016.09.004

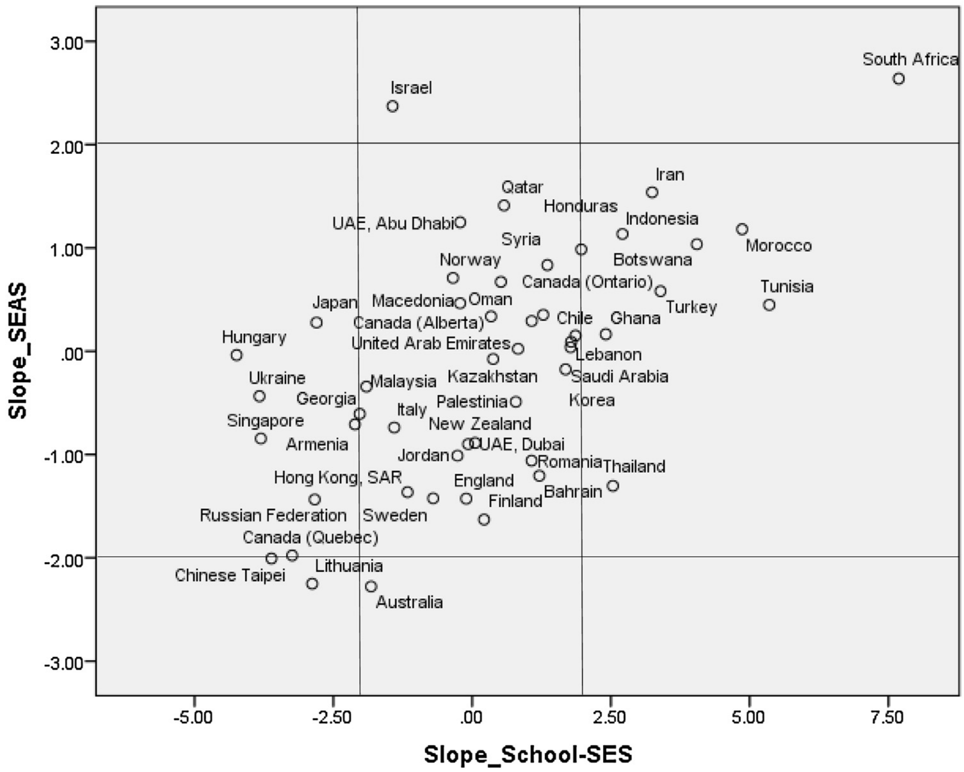


**Fig. 3.** Plot of Slope_SEAS with Slope_School-SES.
*Note*: Slope_SEAS = t-value for the regression of the within-school random slope on school-level latent variable representing School Emphasis on Academic Success; Slope_School-SES = t-value for the regression of the within-school random slope on school SES.

HDI was significantly negatively correlated with Slope_School-SES, Slope_Order and Slope_SEAS and it was negatively correlated with Slope_Hours and Slope_InQua albeit not significantly so (Table 1). This result implies that countries with a higher level of HDI tend to be more compensatory than countries with a lower HDI.

### *3.3. Correlates of level and dispersion of mathematics achievement*

The third research question asked how school differences in equity relate to level and dispersion of mathematics achievement across educational systems. We have identified four possible indicators of equity at the educational system level: the ICC for between-school differences in achievement, Math_SD, the between-school regression of achievement on School-SES (Int_School-SES) and the moderator Slope_School-SES. The first three measures are well-established indicators of equity, while the last one is not.

Int_School-SES only correlated significantly with one of the other variables in the matrix, namely the ICC (Table 1). With the exception of the correlation with Int_School-SES (0.63) the ICC only was significantly correlated with Slope_mean (−0.56). Thus, educational systems with large achievement differences between schools tend to have lower within-school slopes of achievement on SES. One possible explanation for this is that the large school differences are caused by organizational differentiation according to which students are tracked into different schools on the basis of their level of previous achievement. Non-tracked, comprehensive, educational systems would in contrast keep most of the differences between students within schools, allowing for steeper within-school regressions of achievement on SES. However, the fact that neither the ICC, nor Int_School-SES had any significant correlation with level or dispersion of achievement, or with any of the moderators, suggests that these two measures do not capture aspects of equity related to achievement differences at the educational system level.

The school variables most strongly related to mean achievement at the educational system level were Slope_School-SES (-0.64) and Slope_var (-0.52). These two variables were quite highly correlated (.52), and when both variables were entered in a regression model only Slope_School-SES contributed significantly to the prediction of achievement. We therefore concentrate on this variable, and a plot of its relation with mean achievement is shown in Fig. 4.

No educational system with a significant positive Slope_School-SES achieved above the international mean of 500, and with the exception of Korea, the high performing East Asian countries had, as we have already observed, negative values for Slope_School-SES. All educational systems with a significant positive Slope_ School-SES had means below 450. This indicates that high performing countries manage to reduce the importance of students' individual SES for achievement by creating a low relation with School-SES.

The standard deviation and mean of mathematics achievement correlated weakly (−0.24) and only few variables from the within-school regression model correlated with the standard deviation of achievement (see Table 1). The regression of Slope on the moderator safe and orderly school climate (Slope_Order) had the strongest correlation with Math_SD ($r = 0.35$, $p < 0.013$). This result indicates that educational systems with small dispersions of achievement manage to promote equity by means of safe and orderly school climates. Fig. 5 presents the plot of Math_SD with Slope_Order.

The three Canadian educational systems, the three Scandinavian systems and Slovenia had the lowest observed values on Math_SD. These educational systems also had negative values for Slope_Order, which for Canada (Quebec), Norway and Sweden

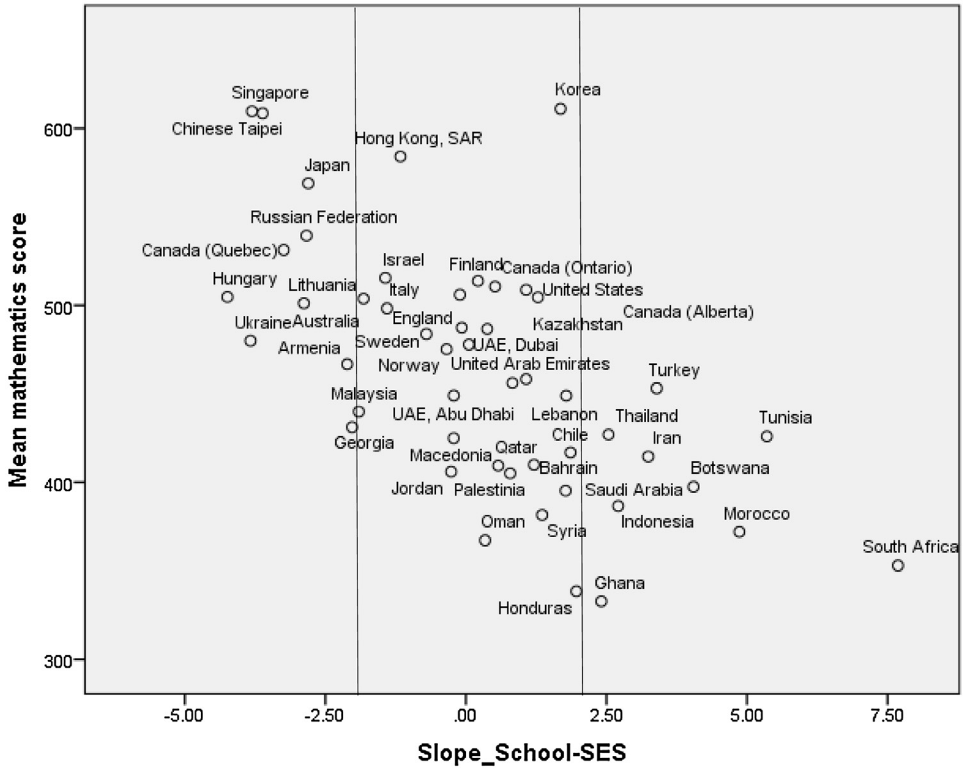


**Fig. 4.** Plot of mean mathematics achievement with Slope_School-SES.
*Note*: Mean mathematics score = Mean of TIMSS 2011 mathematics achievement (first plausible value); Slope_School-SES = t-value for the regression of the random slope on school SES.

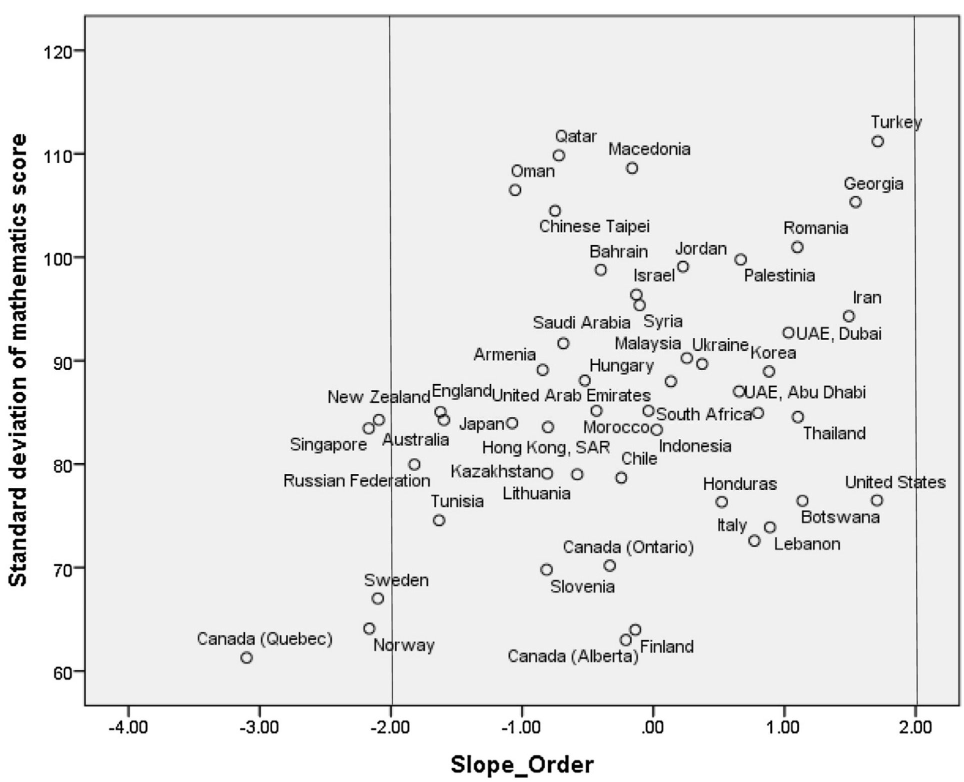


**Fig. 5.** Plot of standard deviation of mathematics score with Slope_Order.
*Note*: Standard deviation of mathematics score = Standard deviation of TIMSS 2011 mathematics achievement (first plausible value); Slope_Order = t-value for the regression of the random slope on school-level latent variable representing Safe and Orderly school.

Please cite this article in press as: J.-E. Gustafsson, et al., 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 (2016), http://dx.doi.org/10.1016/j.stueduc.2016.09.004

were statistically significant. Thus, for these educational systems, higher levels of a safe and orderly school climate were related to a less steep within-school relation between achievement and student SES.

Ample research has demonstrated large differences in level of achievement between educational systems as a function of the level of development in economic and other terms (Mullis & Martin, 2007). However, less is known about levels and mechanisms of equity which makes it interesting to investigate relations between the parameters of the model and the HDI more closely.

As seen in Table 1 there was a very strong correlation between the HDI and Math_mean (0.79) and also a substantial negative correlation between HDI and Math_SD (−0.37). Thus, higher development was associated with higher achievement and a smaller dispersion of achievement. Table 1 also shows that there were relatively strong negative correlations between the HDI on the one hand and the moderators reflecting school climate and school-SES (Slope_School-SES, Slope_Order and Slope_SEAS) on the other hand. These results indicate that the more highly developed countries are capable of reducing the relation between within-school achievement and student SES through creating safe and orderly school climates, emphasizing academic success, and minimizing effects of school SES.

The strong correlations between the HDI and the slope regression parameters raise the question whether the latter make any unique contribution to the dependent variables Math_mean and Math_SD after controlling for HDI. In order to investigate this question a series of regression analyses were ran in which the HDI was combined with one of the five Slope regression parameters at a time. For Math_mean these analyses showed that Slope_ School-SES had a significant contribution over and above HDI, the partial regression coefficient being $\beta = -0.29$ ($t = -2.95$, $p < 0.005$). The other four Slope regressions did not contribute to the prediction of mathematics achievement over and above HDI. For Math_SD, Slope_Order contributed significantly to the prediction, the partial regression coefficient being $\beta = 0.31$ ($t = 2.13$, $p < 0.039$), but no other Slope regression predictor did.

## 4. Discussion

The study asked three research questions: (1) to what extent do within-school regressions of achievement on SES differ across schools and educational systems; (2) to what extent can slope differences be accounted for by school characteristics reflecting quality and quantity of instruction, school climate, and school SES; and (3) how do indicators of equity relate to level and dispersion of mathematics achievement across educational systems? These three questions are discussed below.

### *4.1. To what extent do within-School SES and achievement relations differ across schools?*

The estimates of slope variance for the within-school SES-achievement relation varied across educational systems, there being a clear tendency for educational systems in developing countries to have larger slope variance. However, some exceptions to this pattern were the relatively large slope variances in New Zealand and England.

A large slope variance is indicative of heterogeneous schools, in which students are offered different opportunities to achieve well as a function of SES. However, the measure of slope variance by itself is difficult to interpret, given that no standards have been established for how to interpret it, and because there are many different factors that may influence the steepness of the slope. We therefore need to find determinants of the slope variability.

We also observed lower slope means in educational systems with large slope variance. At first glance it is unexpected that a lower within-school relation between achievement and student SES is indicative of inequity of schooling. However, to the extent that students are selected and self-sorted into different schools on the basis of SES and previous achievement we can expect the within-school regression slopes to become lower (Burger, 2016). With such sorting of students we may also expect between-school differences in level of achievement to become larger, as well as steeper between-school regressions of achievement on school-SES. This pattern of results was also seen, with a significant negative correlation between the ICC and the mean of the within-school regression slope, and a significant positive correlation between the ICC and the between-school regression of achievement on school-SES. The between-school regression coefficient also correlated negatively with the mean within-school regression coefficient (−0.21), albeit not significantly.

These results thus show that a higher mean within-school regression of achievement on SES is associated with smaller between-school differences in achievement, which is generally interpreted as an indication of a more equitable school system. We return below to a more extended discussion of how to interpret different indicators of equity at the school system level.

### *4.2. What accounts for slope variability?*

Based on our review of the literature we hypothesized that school characteristics reflecting quality and quantity of instruction, school climate, and school SES are potential determinants of slope variability across schools. The empirical results provided support for these expectations even though we only found significant moderation in four educational systems for quantity of instruction and in five systems for quality of instruction. For quantity of instruction all coefficients were negative, while for quality of instruction the coefficient was negative in three cases and positive in two cases, These results thus suggest that the effect of quantity of instruction generally was compensatory with respect to SES, which finding is in line with much previous research (e.g., Burger, 2016; Wenglinsky, 1998).

Quality of instruction, in contrast, generated both compensatory and anti-compensatory effects. A compensatory effect may be hypothesized for school systems with policies emphasizing equity, while an anti-compensatory effect may be expected in elitist educational systems where high-quality teaching is allocated to high-achieving student. These hypotheses may be investigated more closely by comparing the organisation of teaching of mathematics in the compensatory systems of Singapore, Hong Kong and Chinese Taipei on the one hand and the anti-compensatory systems of Hungary and Armenia on the other hand.

For school-climate the moderation was significant in six educational systems, and in four cases the coefficients was negative, indicating a compensatory effect. All these educational systems had a high level of HDI, which suggests that the compensatory effect of school climate varies as a function of level of human development. Most previous research which has found a similar pattern of results has also been conducted on high HDI educational systems (e.g. Lee & Smith, 1999; Liu et al., 2015; Thapa et al., 2013).

Our results showed school-SES to be the most powerful predictor of slope variability. The group of educational systems where school-SES was anti-compensatory was primarily composed of developing countries, while the group of systems where school-SES was compensatory included East Asian and Eastern European/former Soviet Union educational systems. Only few previous studies have investigated how school-SES moderates the relation between SES and achievement (Van Ewijk & Sleegers, 2010). However, Burger (2016) found that educational systems with organizational differentiation had steeper relations between SES and achievement than comprehensive educational systems. Thus, if the developing countries to a larger extent use organizational differentiation these results are in agreement with the present ones.

Several different mechanisms may account for the anti-compensatory effect of school-SES. One is peer effects, and positive peer effects may be expected when school SES is high, while negative peer effects may be expected when school SES is low (Van Ewijk & Sleegers, 2010). However, the evidence concerning peer effects is somewhat inconsistent, and even though peer effects may be a partial explanation, other factors are likely to be involved as well. The review of the literature identified another type of compositional effect, namely that quality of instruction and school climate tends to be better in high-SES schools, because students in such schools are better prepared for school-work and are under stronger parental pressure to behave and achieve well. Several studies that we reviewed have demonstrated both student and school SES effects to be mediated via quantity and quality of instruction, low SES students typically being provided with less instruction and instruction of lower quality (Rjosk et.al, 2014; Schmidt et al., 2015; Willms, 2010). Given that previous research also has found resources to give stronger positive effects for low-SES students than high-SES students (e.g., Wenglinsky, 1998), distribution of quality and quantity of instruction which disfavours low-SES students will have even stronger negative effects than is indicated by an additive model.

To summarize, we thus hypothesize that the anti-compensatory effect of school-SES in certain educational systems is due to a higher level of quality and quantity of instruction in high SES schools, a part of which may be due to compositional SES effects, and a part of which may be due to unequal distribution of access to good education across different social groups.

One possible reason for why school-SES has a compensatory effect in certain school systems is that there is less of organizational differentiation of students across different schools in these systems. The more heterogeneous school composition of students in comprehensive school systems would imply that there is less room for school-SES to exert compositional effects and to relate to unequal access to good education. The results showed that anti-compensatory educational systems tend to have lower within-school slopes of achievement on student SES than compensatory educational systems, and we have already suggested that this is because the latter tend to be comprehensive schools with large individual differences in both SES and achievement within schools.

However, many comprehensive educational systems, such as the Scandinavian ones, were not classified into the compensatory category, because their Slope_School-SES estimates were close to zero rather than negative, thus indicating neither compensatory, nor anti-compensatory schooling. Comprehensive schooling is thus not a sufficient condition to achieve compensatory effects with respect to SES. Indeed, a recent study investigating the mediating role of teacher quality for the relation between SES and achievement across the Scandinavian countries indicated a positive association between school-SES and teacher quality (Nilsen, Kaarstein, & Gustafsson, 2016). It is hence reasonable to expect that even within these countries there may be unequal access to good education. Assuming that it is possible to achieve perfect equity in access to good education across all social groups, we could expect a weak compensatory effect, given that the level of achievement would be more strongly affected for low-SES students. However, to achieve a strong compensatory effect it would probably be necessary to use compensatory resource allocation strategies, and provide low SES schools with the more qualified teachers and principals, and make sure that they offer a larger quantity of education. It is an interesting task for further research to investigate if support can be obtained for this hypothesis.

### 4.3. Relations with level of mathematics achievement

Our results showed that the coefficient for Slope on School-SES correlated highly negatively with mathematics achievement at the educational system level, and it even had predictive power over and above HDI. In other words, we identified a relation between equity and achievement. This finding is in line with results reported by Kyriakides, Charalambous, Charalambous, and Dimosthenous (2016) who in reanalyses of PISA data found equity and achievement to be positively related for schools and countries. However, our results also showed that other commonly used indicators of equity of school systems, such as the ICC and the between-school regression of achievement on SES only had weak relations with mean achievement at the educational system level.

Even though we should be careful not to interpret relations in these cross-sectional data as representing causal effects, the strong correlation between math achievement and the regression coefficient for Slope on School-SES suggests the following hypothesis: What matters for level of achievement of an educational system is not the amount of differences between schools in level of achievement or the school-level relations between SES and achievement, but how school-level SES relates to the within-school relations between student SES and achievement. When this relation is low or negative, the educational system is more likely to reach higher levels of achievement than when it is positive. Thus, educational systems perform better when school-SES negates the effect of students' individual SES, so that compensatory effects are obtained.

### 4.4. Relations with dispersion of mathematics achievement

It might have been expected that the coefficient for Slope on School-SES would be negatively correlated with the standard deviation of achievement but these two variables were uncorrelated. The reason for this was that the larger within-school differences in achievement associated with negative values of the Slope_School-SES coefficient caused the overall standard deviation of achievement to be high.

Negative coefficients were, unexpectedly, found for the influence of a safe and orderly climate on the Slope (Slope_Order) for the Scandinavian and Canadian educational systems, which had the lowest standard deviations. Thus, for these educational systems higher levels of safe and orderly climates were related to a less steep within-school relation between achievement and student SES. One possible interpretation of this is that these educational systems have managed to find ways of implementing safety and order in schools which support the motivation and achievement of low SES students. Since the low-SES students tend to be low achievers this would also cause the standard deviation of mathematics achievement to be reduced. However, this empirical finding needs to be replicated, and the proposed interpretation needs to be carefully evaluated and tested. Given that the distinction between comprehensive and organizationally differentiated school systems has been shown to be important for understanding the relation between SES and achievement (Burger, 2016), it does seem essential that differences between these two types of school systems are attended to in this research.

### 4.5. Limitations

Even though this study is based on a massive amount of data, it suffers from certain limitations. One limitation is that the measurement of all the school factors was not optimal. Hours, SEAS and Order were based on one, three and ten items in the principal questionnaire, respectively, and given that there was only one response for each school the reliability of these measures could have been higher. InQua and SES were based on the student questionnaire. For InQua, three items were used, and given that each student rated each item, the school-level reliability should have been acceptable. However, some of the aspects of teaching asked about were not easy to assess, which in combination with cultural differences in response styles, may have threatened validity. The students were, furthermore, asked to assess their individual teacher, while the results were analysed at the school level. In most instances, only one class was sampled for each school, and given the substantial differences in skills among teachers within schools, this may be a poor basis for generalization to the school. However, in some countries more than one class was sampled from each school, which should provide better estimates at the school level, but which also reduces comparability between countries.

SES was measured with a systematically developed scale based on six items asking about factual information, so particularly at the school-level, this variable is likely to have been reliable and valid. This may be one of the reasons why so many more significant relations with Slope were found for school SES than for the other school characteristics. However, we cannot guarantee that this SES measure has equally good measurement characteristics for all countries, which should be investigated in further research. Even though there are reasons to try to improve the reliability and validity of the measures of school characteristics in future research it may be noted that they all provided interpretable patterns of results, albeit with a small number of significant findings.

As was mentioned above, some countries sampled only one class per school, while other countries sampled more than one class per school, if available. This implies that the amount of student heterogeneity of the single class will be taken to reflect the student heterogeneity of the school, which may be incorrect, and particularly so when classes are formed through within-school ability grouping. The consequences that this may have need to be investigated in further research.

Yet another limitation of the study is the cross-sectional design, which precludes causal inference. While it is unlikely that it will be possible to implement individual-level longitudinal designs in future large-scale comparative studies, it may be noted that the trend design of most of these studies allows for longitudinal analyses at the educational system level with for example difference-in-differences techniques, which provide a stronger basis for causal inference (Gustafsson, 2013). Such approaches could be applied in future research.

Yet another limitation is that the results need replication, given that there is very little previous research on the relations investigated here. However, many datasets are available which allow replication and extension of the results, so this should be easy to accomplish.

## 5. Conclusions

The present study found that some educational systems function in a compensatory manner with respect to the association between SES and achievement thus improving equity, while others rather are anti-compensatory. The findings also showed that the compensatory systems tend to have higher levels of achievement. A partial explanation for this pattern is that the more highly developed countries are more capable of reducing the relation between within-school achievement and student SES through healthy school climates where order and safety prevail and where there is high priority for academic success, and higher instructional quality. These educational systems also avoid peer effects which increase SES effects and they allocate more and better instruction to low-SES students.

Our results also indicated that the direction and strength of the relation between school-SES and the within-school association between student SES and achievement is a powerful measure of equity with better characteristics than other measures of equity such as the ICC, the between-school regression of achievement on SES or the standard deviation of achievement. Thus, future research should be directed to further investigations of the properties of this measure and of the mechanisms that it reflects.

## Appendix A. Descriptive statistics including number of students and schools, level and dispersion of math achievement, ICC, HDI and descriptives of SES for 50 educational systems in TIMSS 2011 Grade 8

| Country | The sample | | Mathematics Achievement | | | Human Development Index | SES | | |
|---|---|---|---|---|---|---|---|---|---|
| | No of schools | No of students | Math mean | Math SD | ICC | HDI | SES Mean | SES SD Student | SES SD School |
| Armenia | 153 | 5846 | 467 | 89 | 0.22 | 0.73 | 10.88 | 1.81 | 0.82 |
| Australia | 277 | 7556 | 504 | 84 | 0.52 | 0.93 | 11.23 | 1.64 | 0.85 |
| Bahrain | 95 | 4640 | 410 | 99 | 0.42 | 0.81 | 10.16 | 1.73 | 0.65 |
| Botswana | 150 | 5400 | 397 | 76 | 0.19 | 0.68 | 8.36 | 1.99 | 0.69 |
| Canada (Alberta) | 145 | 4799 | 505 | 63 | 0.16 | 0.90 | 11.53 | 1.55 | 0.57 |
| Canada (Ontario) | 143 | 4752 | 511 | 70 | 0.16 | 0.90 | 11.46 | 1.55 | 0.64 |
| Canada (Quebec) | 189 | 6149 | 531 | 61 | 0.35 | 0.90 | 11.16 | 1.46 | 0.66 |
| Chile | 193 | 5835 | 417 | 79 | 0.52 | 0.82 | 9.75 | 1.77 | 1.10 |
| Chinese Taipei | 150 | 5042 | 609 | 104 | 0.22 | Miss | 10.48 | 1.79 | 0.83 |
| England | 118 | 3842 | 506 | 84 | 0.67 | 0.89 | 10.85 | 1.68 | 0.87 |
| Finland | 145 | 4266 | 514 | 64 | 0.14 | 0.88 | 11.32 | 1.53 | 0.49 |
| Georgia | 172 | 4563 | 431 | 105 | 0.30 | 0.74 | 10.62 | 1.88 | 1.05 |
| Ghana | 161 | 7323 | 333 | 85 | 0.39 | 0.57 | 7.75 | 2.03 | 1.10 |
| Honduras | 155 | 4418 | 338 | 76 | 0.33 | 0.62 | 8.38 | 2.14 | 1.18 |
| Hong Kong, SAR | 117 | 4015 | 584 | 84 | 0.63 | 0.89 | 9.95 | 1.79 | 0.95 |
| Hungary | 146 | 5178 | 505 | 88 | 0.33 | 0.82 | 10.85 | 1.82 | 1.00 |
| Indonesia | 153 | 5795 | 387 | 83 | 0.42 | 0.68 | 8.26 | 1.68 | 0.81 |
| Iran | 238 | 6029 | 415 | 94 | 0.44 | 0.75 | 8.54 | 2.48 | 1.66 |
| Israel | 151 | 4699 | 516 | 96 | 0.39 | 0.89 | 10.93 | 1.82 | 0.97 |
| Italy | 197 | 3979 | 498 | 73 | 0.24 | 0.87 | 10.38 | 1.75 | 0.80 |
| Japan | 138 | 4414 | 569 | 84 | 0.13 | 0.89 | 10.86 | 1.65 | 0.62 |
| Jordan | 230 | 7694 | 406 | 99 | 0.27 | 0.74 | 9.55 | 1.90 | 0.81 |
| Kazakhstan | 147 | 4390 | 487 | 79 | 0.45 | 0.76 | 10.11 | 1.69 | 0.96 |

Please cite this article in press as: J.-E. Gustafsson, et al., 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 (2016), http://dx.doi.org/10.1016/j.stueduc.2016.09.004

(*Continued*)

| Country | The sample | | Mathematics Achievement | | | Human Development Index | SES | | |
|---|---|---|---|---|---|---|---|---|---|
| | No of schools | No of students | Math mean | Math SD | ICC | HDI | SES Mean | SES SD Student | SES SD School |
| Korea | 150 | 5166 | 611 | 89 | 0.09 | 0.89 | 11.47 | 1.81 | 0.70 |
| Lebanon | 147 | 3974 | 449 | 74 | 0.44 | 0.76 | 9.36 | 2.06 | 1.15 |
| Lithuania | 141 | 4747 | 501 | 79 | 0.23 | 0.83 | 10.52 | 1.62 | 0.77 |
| Macedonia | 150 | 4062 | 425 | 109 | 0.38 | 0.73 | 9.97 | 1.69 | 0.88 |
| Malaysia | 180 | 5733 | 440 | 90 | 0.67 | 0.77 | 9.07 | 1.84 | 1.02 |
| Morocco | 279 | 8985 | 372 | 85 | 0.35 | 0.61 | 7.92 | 2.37 | 1.27 |
| New Zealand | 158 | 5336 | 487 | 85 | 0.37 | 0.91 | 10.99 | 1.75 | 0.77 |
| Norway | 134 | 3862 | 475 | 64 | 0.13 | 0.94 | 11.68 | 1.61 | 0.54 |
| Oman | 323 | 9542 | 367 | 106 | 0.32 | 0.78 | 9.02 | 2.16 | 0.97 |
| Palestine | 201 | 7812 | 405 | 100 | 0.23 | 0.68 | 9.17 | 2.01 | 0.80 |
| Qatar | 109 | 4422 | 409 | 110 | 0.48 | Miss | 10.77 | 1.80 | 0.84 |
| Romania | 147 | 5523 | 458 | 101 | 0.35 | 0.78 | 9.94 | 1.85 | 1.18 |
| Russian Fed | 210 | 4893 | 539 | 80 | 0.41 | 0.78 | 10.96 | 1.64 | 0.90 |
| Saudi Arabia | 153 | 4344 | 395 | 92 | 0.34 | 0.83 | 9.35 | 2.13 | 1.13 |
| Singapore | 165 | 5927 | 610 | 83 | 0.44 | 0.90 | 10.36 | 1.77 | 0.79 |
| Slovenia | 186 | 4415 | 505 | 70 | 0.10 | 0.87 | 10.97 | 1.44 | 0.46 |
| South Africa | 285 | 11966 | 353 | 85 | 0.64 | 0.65 | 8.67 | 1.97 | 0.99 |
| Sweden | 153 | 5568 | 484 | 67 | 0.14 | 0.90 | 11.41 | 1.66 | 0.64 |
| Syria | 148 | 4413 | 382 | 95 | 0.31 | 0.66 | 8.65 | 2.11 | 1.16 |
| Thailand | 172 | 6124 | 427 | 85 | 0.56 | 0.72 | 8.40 | 2.01 | 1.21 |
| Tunisia | 207 | 5128 | 426 | 75 | 0.32 | 0.72 | 8.91 | 2.04 | 1.07 |
| Turkey | 239 | 6928 | 453 | 111 | 0.30 | 0.76 | 8.28 | 2.33 | 1.55 |
| UAE, Abu Dhabi | 166 | 4373 | 449 | 87 | 0.37 | 0.83 | 10.38 | 1.78 | 0.83 |
| UAE, Dubai | 130 | 5571 | 478 | 93 | 0.47 | 0.83 | 10.70 | 1.76 | 0.88 |
| Ukraine | 148 | 3378 | 480 | 90 | 0.25 | 0.73 | 10.54 | 1.61 | 0.85 |
| UAE | 458 | 14089 | 456 | 88 | 0.40 | 0.83 | 10.39 | 1.78 | 0.83 |
| United States | 501 | 10477 | 509 | 76 | 0.58 | 0.91 | 10.98 | 1.87 | 1.10 |

*Note*: Math mean = Mean of TIMSS 2011 mathematics achievement (first plausible value); Math SD = standard deviation of TIMSS 2011 mathematics achievement (first plausible value); ICC = Intraclass correlation of between-school differences with respect to TIMSS 2011 mathematics achievement; HDI = Human Development Index 2012, as defined by UNHDP (2014); SES mean = mean of socio-economic status; SES SD Student = within-school pooled standard deviation of SES; SES SD School = school-level standard deviation of SES.

## Appendix B. Estimates from random slopes models for 50 educational systems in TIMSS 2011 Grade 8

| Country | The Slope | | Moderator t-values | | | | | Int_School-SES |
|---|---|---|---|---|---|---|---|---|
| | Slope mean | Slope var | Slope_ InQua | Slope_ Hours | Slope_ SEAS | Slope_ Order | Slope_ School SES | |
| Armenia | 0.28* | 0.002 | 2.31* | 0.17 | −0.71 | −0.84 | −2.11* | 1.27* |
| Australia | 0.24* | 0.002 | −0.93 | −1.39 | −2.28* | −2.09* | −1.82 | 2.25* |
| Bahrain | 0.25* | 0.006 | −0.61 | −1.03 | −1.21 | −0.4 | 1.21 | 2.70* |
| Botswana | −0.04* | 0.01* | −0.56 | −0.04 | 1.04 | 1.14 | 4.05* | 1.82* |
| Canada (Alberta) | 0.25* | 0.002 | −0.39 | −1.66 | 0.35 | −0.21 | 1.28 | 1.00* |
| Canada (Ontario) | 0.27* | 0 | −1.63 | −1.39 | 0.67 | −0.33 | 0.52 | 1.07* |
| Canada (Quebec) | 0.24* | 0.001 | 0.48 | −0.29 | −1.98* | −3.1 | −3.24* | 1.18* |
| Chile | 0.13* | 0 | −0.66 | −1.35 | 0.15 | −0.24 | 1.86 | 1.69* |
| Chinese Taipei | 0.47* | 0.001 | −2.54* | −2.65* | −2.01* | −0.75 | −3.62* | 1.65* |
| England | 0.22* | 0.007 | −0.47 | −0.57 | −1.43 | −1.6 | −0.11 | 2.81* |
| Finland | 0.32* | 0.001 | −0.37 | −1.46 | −1.63 | −0.14 | 0.22 | 1.04* |
| Georgia | 0.37* | 0 | −1.42 | −2.02* | −0.61 | 1.54 | −2.03* | 1.39* |
| Ghana | −0.09* | 0.012* | 1.36 | 1.19 | 0.16 | −0.04 | 2.41* | 0.87* |
| Honduras | 0.03 | 0.004 | 0.57 | 1.87 | 0.99 | 0.52 | 1.97* | 1.21* |
| Hong Kong | 0.04* | 0.002 | −2.58* | −0.52 | −1.37 | −0.8 | −1.17 | 2.14* |
| Hungary | 0.46* | 0.003 | 2.36* | 0.67 | −0.04 | −0.52 | −4.24* | 1.46* |
| Indonesia | −0.01 | 0.006 | 0.1 | 0.38 | 1.14 | 0.03 | 2.71* | 1.31* |
| Iran | 0.19* | 0.003 | −0.51 | 0.56 | 1.54 | 1.49 | 3.24* | 1.26* |
| Israel | 0.30* | 0 | 1.58 | −0.19 | 2.37* | −0.13 | −1.43 | 2.11* |
| Italy | 0.31* | 0.001 | −0.63 | −1.61 | −0.74 | 0.77 | −1.4 | 0.88* |
| Japan | 0.41* | 0 | −1.06 | −0.89 | 0.28 | −1.08 | −2.8* | 1.55* |
| Jordan | 0.29* | 0 | 0.27 | 0.1 | −1.01 | 0.23 | −0.26 | 1.59* |
| Kazakhstan | 0.16* | 0.009* | 1.73 | −0.71 | −0.08 | −0.81 | 0.38 | 1.05* |
| Korea | 0.57* | 0 | 0.64 | −0.61 | −0.18 | 0.88 | 1.68 | 0.85* |
| Lebanon | 0.07 | 0.003 | 0.2 | 0.39 | 0.09 | 0.89 | 1.78 | 1.55* |
| Lithuania | Miss | Miss | 0.52 | −1.3 | −2.25* | −0.58 | −2.88* | Miss |
| Macedonia | 0.4* | 0.013* | −1.15 | −0.5 | 0.47 | −0.16 | −0.22 | 2.55* |
| Malaysia | 0.07* | 0.002 | 0.93 | 0.52 | −0.34 | 0.26 | −1.91 | 2.53* |
| Morocco | 0.14* | 0.009* | 1.56 | −0.76 | 1.18 | −0.43 | 4.86* | 1.32* |
| New Zealand | 0.35* | 0.01* | −0.67 | −1.64 | −0.9 | −1.62 | −0.07 | 2.54* |
| Norway | 0.37* | 0 | −0.26 | 1.61 | 0.71 | −2.17* | −0.35 | 1.09* |
| Oman | 0.34* | 0.011* | −1.32 | −2.96* | 0.34 | −1.05 | 0.34 | 1.23* |
| Palestinia | 0.33* | 0.008 | 0.03 | 0.12 | −0.49 | 0.67 | 0.79 | 0.70* |
| Qatar | 0.24* | 0.001 | 1.3 | −0.35 | 1.41 | −0.72 | 0.57 | 3.05* |
| Romania | 0.50* | 0.001 | −1.23 | 0.9 | −1.06 | 1.1 | 1.07 | 1.14* |

Please cite this article in press as: J.-E. Gustafsson, et al., 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 (2016), http://dx.doi.org/10.1016/j.stueduc.2016.09.004

(*Continued*)

| Country | The Slope | | Moderator t-values | | | | | Int_School-SES |
|---|---|---|---|---|---|---|---|---|
| | Slope mean | Slope var | Slope_ InQua | Slope_ Hours | Slope_ SEAS | Slope_ Order | Slope_ School SES | |
| Russian Fed | 0.18* | 0 | −0.18 | 0.22 | −1.44 | −1.82 | −2.84* | 1.39* |
| Saudi Arabia | 0.15* | 0.009 | 0.82 | −1.09 | 0.04 | −0.69 | 1.77 | 0.94* |
| Singapore | 0.17* | 0.001 | −3.04* | 1.01 | −0.85 | −2.17* | −3.81* | 2.46* |
| Slovenia | 0.43* | 0.001 | 1.51 | Miss | −1.84 | −0.81 | Miss | 1.26* |
| South Africa | 0.01 | 0.01* | −0.89 | −1.38 | 2.64* | 0.8 | 7.69* | 2.41* |
| Sweden | 0.35* | 0 | 0.42 | 0.09 | −1.42 | −2.1* | −0.7 | 1.22* |
| Syria | 0.07* | 0.008 | −1.33 | −0.63 | 0.83 | −0.1 | 1.36 | 0.60* |
| Thailand | 0.08* | 0.002 | −1.67 | −2.13* | −1.3 | 1.1 | 2.54* | 1.42* |
| Tunisia | 0.14* | 0.013* | −0.22 | 1.32 | 0.45 | −1.63 | 5.35* | 1.02* |
| Turkey | 0.36* | 0.015* | 0.41 | −0.9 | 0.58 | 1.71 | 3.39* | 1.08* |
| UAE, Abu Dhabi | 0.19* | 0.016* | 0.29 | −0.14 | 1.25 | 0.65 | −0.22 | 1.74* |
| UAE, Dubai | 0.21* | 0.005 | 0.46 | 0.13 | −0.89 | 1.03 | 0.05 | 1.97* |
| Ukraine | 0.39* | 0 | 0.24 | 0.38 | −0.44 | 0.37 | −3.83* | 1.58* |
| UAE | 0.20* | 0.007* | 0.89 | 0.54 | 0.02 | 0.13 | 0.83 | 1.69* |
| United States | 0.12* | 0 | −0.23 | −1.22 | 0.29 | 1.71 | 1.07 | 1.70* |

*Note*: UAE = United Arabian Emirates; Slope mean = Mean of within-school regression slopes for student math score on SES; Slope var = Variance of within-school regression slopes for student math score on SES; Slope_InQua = t-value for the regression of the within-school Slope on school-level latent variable representing instructional quality; Slope_Hours = t-value for the regression of the within-school Slope on school-level yearly hours of instruction; Slope_SEAS = t-value for the regression of the within-school Slope on school-level latent variable representing School Emphasis on Academic Success; Slope_Order = t-value for the regression of the within-school Slope on school-level latent variable representing Safe and Orderly school; Slope_School-SES = t-value for the regression of the within-school Slope on mean school SES; Int_School-SES = coefficient for the regression of school-level achievement on School-SES.

*Coefficient is significant at the 0.05 level (2-tailed).

## References

Baumert, J., et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*(1), 133–180.
Borman, G., & Dowling, M. (2010). Schools and inequality: A multilevel analysis of Coleman's equality of educational opportunity data. *The Teachers College Record*, *112*(5), 1–2.
Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258).New York: Greenwood.
Brantlinger, E. A. (2003). *Dividing classes: How the middle class negotiates and rationalizes school advantage*. New York: RoutledgeFalmer.
Bryk, A. S., & Raudenbush, S. W. (1988). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model? *American Journal of Education*, *97*(1), 65–108.
Bryk, A., & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. New York, NY: Russell Sage Foundation.
Burger, K. (2016). Intergenerational transmission of education in Europe: Do more comprehensive education systems reduce social gradients in student achievement? *Research in Social Stratification and Mobility*, *44*, 54–67.
Burtless, G. T. (1996). *Does money matter? The effect of school resources on student achievement and adult success*. Brookings Institution Press.
Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*S95–S120.
Ecalle, J., Magnan, A., & Gibert, F. (2006). Class size effects on literacy skills and literacy interest in first grade: A large-scale investigation. *Journal of School Psychology*, *44*, 191–209.
Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Willms, J. D. (2001). Class size and student achievement. *Psychological Science in the Public Interest*, *2*(1), 1–30.
Finn, J. D., & Achilles, C. M. (1999). Tennesse's class size study: Findings implications and misconceptions. *Educational Evaluation and Policy Analysis*, *21*, 97–110.
Goldstein, S. E., Young, A., & Boyd, C. (2008). Relational aggression at school: Associations with school safety and social climate. *Journal of Youth and Adolescence*, *37*, 641–654.
Good, T. L., Wiley, C. R., & Florez, I. R. (2009). Effective teaching: An emerging synthesis. *International handbook of research on teachers and teaching*. US: Springer803–816.
Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, *66*, 361–396.
Gregory, A., Cornell, D., & Fan, X. (2012). Teacher safety and authoritative school climate in high schools. *American Journal of Education*, *118*(4), 401–425.
Gustafsson, J. E. (2013). Causal inference in educational effectiveness research: a comparison of three methods to investigate effects of homework on student achievement. *School Effectiveness and School Improvement*, *24*(3), 275–295.
Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher*, *18*(4), 45–65.
Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students'' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404).Charlotte, NC: Information Age Publishing.
Hoy, W. K., Tarter, C. J., & Hoy, A. W. (2006). Academic optimism of schools: A force for student achievement. *American Educational Research Journal*, *43*(3), 425–446.
IDB Analyzer (Version 3.1) [Computer software and manual]. (2015). Hamburg, Germany: International Association for the Evaluation of Educational Achievement.
Johnson, B., & Stevens, J. J. (2006). Student achievement and elementary teachers' perceptions of school climate. *Learning Environments Research*, *9*, 111–122.
Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study: Investigating effects ofTeaching and learning in swiss and german classrooms. In T. Janik, & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160).Münster: Waxmann Verlag.
Krueger, A. B. (2003). Economic considerations and class size. *Economic Journal*, *113*, F34–F63.
Kyriakides, L., Creemers, B., Antoniou, P., & Demetriou, D. (2010). A synthesis of studies searching for school factors: Implications for theory and research. *British Educational Research Journal*, *36*(5), 807–830.
Kyriakides, L., Charalambous, E., Charalambous, C. Y., & Dimosthenous, A. (2016). Using international comparative studies to investigate the relationship between quality and equity in education. *Paper presented at the annual meeting of the american educational research association* (pp. 2016.
Lee, V. E., & Smith, J. B. (1999). Social support and achievement for young adolescents in Chicago: The role of school academic press? *American Educational Research Journal*, *36*(4), 907–945.
Liberatos, P., Link, B. G., & Kelsey, J. L. (1987). The measurement of social class in epidemiology. *Epidemiologic Reviews*, *10*, 87–121.
Liu, H., Van Damme, J., Gielen, S., & Van Den Noortgate, W. (2015). School processes mediate school compositional effects: Model specification and estimation. *British Educational Research Journal*, *41*(3), 423–447.
Marks, G. N., Cresswell, J., & Ainley, J. (2006). Explaining socioeconomic inequalities in student achievement: The role of home and school factors? *Educational Research and Evaluation*, *12*(2), 105–128.
Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J., Abduljabbar, A. S., et al. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist*, *47*(2), 106–124.
Martin, M. O., Mullis, I. V., Foy, P., & Arora, A. (2012). Creating and interpreting the TIMSS and PIRLS 2011 context questionnaire scales. In M. O. Martin, & I. V. S. Mullis (Eds.), *Methods and procedures in TIMSS and PIRLS 2011*Chestnut Hill, MA: TIMSS & PIRLS International Study Center Boston College.
Martin, M. O., Foy, P., Mullis, I. V. S., & O'Dwyer, L. M. (2013). Effective schools in reading, mathematics, and science at the fourth grade. In M. O. Martin, & I. V. S. Mullis (Eds.), *TIMSS and PIRLS 2011: Relationships among reading, mathematics, and science achievement at the fourth grade—Implications for early learning*Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
Maxwell, S. E., Cole, D. A., & Mitchell, M. A. (2011). Bias in cross-sectional analyses of longitudinal mediation: Partial and complete mediation under an autoregressive model. *Multivariate Behavioral Research*, *46*(5), 816–841.

Please cite this article in press as: J.-E. Gustafsson, et al., 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 (2016), http://dx.doi.org/10.1016/j.stueduc.2016.09.004

McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist*, *53*(2), 185–204.
Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology*, *110*(1), 40.
Mullis, I. V. S., & Martin, M. O. (2007). TIMSS in perspective: Lessons learned from IEA's four decades of international mathematics assessments. In T. Loveless (Ed.), *Lessons learned: What international assessments tell us about math achievement* (pp. 9–36).Brookings Institution Press.
Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 International Results in Mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
Muthén, L., & Muthén, B. (1998). *Mplus user's guide*, 7th ed. Los Angeles, CA: Muthén n.
Nilsen, T., & Gustafsson, J. E. (2014). School emphasis on academic success: exploring changes in science performance in Norway between 2007 and 2011 employing two-level SEM. *Educational Research and Evaluation*, *20*(4), 308–327.
Nilsen, T., Kaarstein, H., & Gustafsson, J. E. (2016). Teacher quality mediating and moderating the relation between SES and achievement in the Nordic countries. *Paper presented at the European Association for Research on Learning and Instruction Sig 18 & 23 meeting*.
Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*(3), 237–257.
OECD (2012). *PISA 2009 technical report*. Paris: OECD.
OECD (2013). *PISA 2012 Results: Excellence through equity. Giving every student the chance to succeed (Volume II)*. PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264201132-en.
Oakes, J. (1985). *Keeping track*. New Haven, CT: Yale University Press.
Opdenakker, M. C., & Damme, J. V. (2007). Do school context: Student composition and school leadership affect school practice and outcomes in secondary education? *British Educational Research Journal*, *33*(2), 179–206.
Palardy, G. J.1 (2013). High school socioeconomic segregation and student attainment. *American Educational Research Journal*, *50*(4), 714–754.
Papanastasiou, C. (2008). Factors distinguishing most and least effective schools in terms of reading achievement: A residual approach. *Educational Research and Evaluation*, *14*(6), 539–549.
Perry, L. B., & McConney, A. (2010). Does the SES of the School Matter? An examination of socioeconomic status and student achievement using PISA 2003. *Teachers College Record*, *112*(4), 7–8.
Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, *38*(2), 109–119.
Rjosk, C., Richter, D., Hochweber, J., Lüdtke, O., Klieme, E., & Stanat, P. (2014). Socioeconomic and language minority classroom composition and individual reading achievement: The mediating role of instructional quality. *Learning and Instruction*, *32*, 63–72.
Rumberger, R., & Palardy, G. (2005). Does segregation still matter? The impact of student composition on academic achievement in high school. *The Teachers College Record*, *107*(9), 1999–2045.
Rutter, M., Maughan, B., Mortimore, P., Ouston, J., & Smith, A. (1979). *Fifteen thousand hours: Secondary schools and their effects on children*. Cambridge, MA: Harvard University Press.
Scheerens, J. (Ed.), (2014). *Effectiveness of time investments in education. Insights from a review and meta-analysis*. Springer Verlag, SpringerBriefs in Education.
Scherer, R., & Gustafsson, J. E. (2015). Student assessment of teaching as a source of information about aspects of teaching quality in multiple subject domains: an application of multilevel bifactor structural equation modeling. *Frontiers in Psychology*, *6*(1550) .
Schmidt, W. H., Burroughs, N. A., Zoido, P., & Houang, R. T. (2015). The role of schooling in perpetuating educational inequality: An international perspective. *Educational Researcher*, *44*(7), 371–386.
Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research 1990–2000? *Review of Educational Research*, *75*(3), 417–453.
Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research? *Review of Educational Research*, *83*(3), 357–385.
Thrupp, M. (1999). *Schools making a difference: Let's be realistic! School mix, school effectiveness and the social limits of reform*. Buckingham, Philadelphia: OpenUniversity Press.
Thrupp, M., Lauder, H., & Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research*, *37*(5), 483–504.
UNDP (2014). *Human Development Report 2014. Sustaining Human Progress: Reducing Vulnerabilities and Building Resilience*. New York: United Nations Development Programme.
Uline, C., & Tschannen-Moran, M. (2008). The walls speak: The interplay of quality facilities, school climate, and student achievement. *Journal of Educational Administration*, *46*(1), 55–73.
Van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review*, *5*(2), 134–150.
Van de Werfhorst, H. G., & Mijs, J. J. (2010). Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual Review of Sociology*, *36*, 407–428.
Wang, M. T., & Degol, J. L. (2015). School climate: A review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review*. http://dx.doi.org/10.1007/s10648-015-9319-1.
Wenglinsky, H. (1998). Finance equalization and within-school equity: The relationship between education spending and the social distribution of achievement? *Educational Evaluation and Policy Analysis*, *20*(4), 269–283.
White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, *91*(3), 461–481.
Willms, J. D. (2010). School composition and contextual effects on student outcomes. *The Teachers College Record*, *112*(4), 3–4.
Yang, Y., & Gustafsson, J.-E. (2004). Measuring socioeconomic status at individual and collective Levels. *Educational Research and Evaluation*, *10*(3), 259–288.
Yang, Y. (2003). *Measuring socioeconomic status and its effects at individual and collective levels: A cross-country comparison (Gothenburg Studies in Educational Sciences, 193)*. Gothenburg, Sweden: Acta Universitatis Gothoburgensis.
von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful. *IERI Monograph Series*, *2*, 9–36.

**Jan-Eric Gustafsson** has since 1986 been Professor of Education at the University of Gothenburg. His substantively oriented research focuses on individual differences in cognitive abilities, and on determinants of educational outcomes at individual and system levels. His methodologically oriented research focuses on conceptual and technical issues of measurement within classical and modern test theory, and on the application of latent variable models to both non-nested and nested data.

**Dr Trude Nilsen** is working as a researcher for the Norwegian TIMSS project at the University of Oslo. She has also been working with professional development of teachers and before that she was a high school teacher in mathematics. Her research focuses on instructional quality and school climate within the field of educational effectiveness research. She has a special interest in large-scale international studies and quantitative methodology.

**Dr. Kajsa Yang Hansen's** research lies mainly in the field of sociology of education, with a main focus on educational quality and equity from comparative perspectives. To study and to explain the variation in academic achievement between individuals and schools across countries and over time has always been her primary research interest. Her methodological focus is on analytical techniques for large-scale survey data, such as multi-level analysis, structural equation modelling (SEM) and second generation SEM.