Policy Research Working Paper 9896

What Explains Boys' Educational Underachievement in the Kingdom of Saudi Arabia?

Mahmoud A. A. Elsayed
Aidan Clerkin
Vasiliki Pitsia
Nayyaf Aljabri
Khaleel Al-Harbi

WORLD BANK GROUP
Education Global Practice
January 2022

Policy Research Working Paper 9896

Abstract

This paper examines the factors that are associated with boys' underachievement in mathematics and science in Saudi Arabia, where students attend gender-segregated schools from grade 1 onward, as well as student achievement in these two subjects in grades 4 and 8 more generally. The paper employs data from two recent large-scale assessments of education: Trends in International Mathematics and Science Study 2019 and Saudi Arabia's National Assessment of Learning Outcomes 2018. The results suggest that in grade 4, school climate was more strongly associated with boys' compared with girls' achievement in both mathematics and science, with boys attending schools of poorer school climate having a considerably lower performance compared with girls attending such schools. The findings also indicate that although greater literacy and numeracy readiness was linked with higher science achievement among boys and girls, grade 4 boys tended to benefit more from this readiness than girls. In addition, the results show that student absenteeism in grade 4 is particularly strongly associated with decreases in mathematics achievement among boys. In grade 8, interactions between student gender and students' confidence in science, the degree of schools' emphasis on academic success, and teachers' age are observed. The paper concludes by discussing some of the implications of these findings for educators and policy makers in Saudi Arabia.

This paper is a product of the Education Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at melsayed@worldbank.org.

What Explains Boys' Educational Underachievement in the Kingdom of Saudi Arabia?*

Mahmoud A. A. Elsayed[a], Aidan Clerkin[b], Vasiliki Pitsia[b], Nayyaf Aljabri[c], Khaleel Al-Harbi[c]

[a] The World Bank, Washington DC, USA

[b] Educational Research Centre, Dublin, Ireland

[c] Education and Training Evaluation Commission, Riyadh, Saudi Arabia

Keywords: Student Achievement, Gender Gap, Boys' Underperformance

JEL classification: I20; I21; I24; I28

* This research has been conducted as part of the Technical Cooperation Program between the World Bank and the Education and Training Evaluation Commission, funded by the Ministry of Finance of the Kingdom of Saudi Arabia under the Reimbursable Advisory Services (RAS) framework.

1. Introduction

There is a large and consistent performance gap observed between boys and girls in the Middle East and North Africa (MENA) region. On top of having the second largest proportion of children in learning poverty (i.e., lacking basic skills in reading toward the end of primary school[1]), MENA has the largest gender gap in learning poverty among regions with available data (Gregory et al. 2021; World Bank 2021).[2] This gender gap in student performance, however, varies across MENA countries. The Kingdom of Saudi Arabia (KSA) is among the countries with the largest gender gaps in student achievement in the world. Boys in Saudi Arabia consistently and significantly underperform compared to girls across different grades and subjects. For example, in the Trends in International Mathematics and Science Study (TIMSS) 2019, grade 4 boys in Saudi Arabia scored below girls by approximately 26 points in mathematics and 60 points in science (Mullis et al. 2020). A similar gender gap exists in grade 8, where boys underperform girls by approximately 17 points in mathematics and 47 points in science. Furthermore, in grade 4, 53 percent of boys failed to achieve minimum proficiency in mathematics compared to 45 percent of girls. Similarly, in grade 8, 58 percent of boys did not achieve the lowest international benchmark (400 points) to be considered as minimally proficient in mathematics, compared to 49 percent of girls. Data from the Programme for International Student Assessment (PISA) 2018 also show that almost twice as many 15-year-old boys (65 percent) failed to achieve minimum proficiency in reading, compared to 38 percent of girls (OECD 2019).

Although the magnitude of the gender gap in Saudi Arabia is among the largest in the world, significant differences in achievement between boys and girls are also observed in many other countries (Mullis et al. 2017; 2020). A large body of research has explored the factors contributing to the gender gap in student performance internationally (Autor et al. 2016; 2019; Bertrand and Pan 2013; Buchmann and DiPrete 2006; DiPrete and Buchmann 2013; DiPrete and Jennings 2012; Fortin, Oreopoulos, and Phipps 2015; Jha and Pouezevara 2016; Legewie and DiPrete 2012; OECD 2021). Evidence from this research suggests that social norms, school characteristics, students' social and behavioral skills, and family background are the main factors associated with the achievement gap between boys and girls.

Prior research indicates that, compared to boys, girls tend to have better noncognitive skills, such as self-regulation and persistence, and spend more time doing assignments and homework (Buchmann and DiPrete 2006; Cornwell, Mustard, and Van Parys 2013; DiPrete and Jennings 2012; Downey and Vogt Yuan 2005; OECD 2021). Other studies found that school characteristics such as school quality, disciplinary practices, and school climate (i.e., the institutional norms, practices, structures, values, and relationships underpinning a student's experience of school) can affect boys and girls differently. An important study combined birth records with school administrative data from the US state of Florida to identify the effects of school quality (with school-level gains in mathematics and reading scores used to indicate quality) on the gender achievement gap between opposite-gender siblings who attend the same sets of schools (Autor et al. 2016). This study found that boys benefitted more than girls from studying in higher quality schools. Similarly, another recent study (OECD 2021) based on data from two large-scale international assessments showed that school discipline problems affect boys more negatively than girls. Combining data from the Teaching and Learning International Survey (TALIS) 2018 and PISA 2018, the study found that increases in teachers' perceptions of classroom discipline problems are associated with an increase in the achievement gap between girls and boys (OECD 2021). The results also showed that other school

[1] More formally, learning poverty is defined as "being unable to read and understand a simple text by age 10", whether children are enrolled in school or not (World Bank 2019, 6).

[2] The learning poverty rate is calculated using the results of international student assessments, adjusted for the share of out-of-school children. Overall, 59 percent of children in MENA are not able to read and understand an age-appropriate text by age 10. About two-thirds of boys in MENA (66 percent) lack basic skills in reading compared to 56 percent of girls, resulting in the largest gender gap in learning poverty among all regions.

organizational issues, such as poor learning conditions and organizational problems, exacerbate the gender gap between girls and boys. That is, boys' achievement tends to be negatively impacted by challenging learning conditions to a greater extent compared to girls' achievement. These findings are consistent with previous work on gender gaps in attitudes, showing that girls, in general, tend to report more positive attitudes toward learning (DiPrete and Buchmann 2013). Previous research has also shown that social norms, gender stereotypes, and teacher/school expectations contribute to the performance gap between girls and boys (Stromquist 2007; Page and Jha 2009; Younger and Cobbett 2014; Jha and Pouezevara 2016). In other words, teachers and schools contribute to developing and reinforcing different expectations of appropriate behavior for both boys and girls, which can have the effect of hindering boys' performance.

Building on the existing research, this paper examines the factors that contribute to the gender gap in student performance in Saudi Arabia. In particular, the paper investigates whether a range of student, family, class, and school variables can explain the achievement gap in mathematics and science between boys and girls at grades 4 and 8. The contribution of these variables in predicting overall Saudi student achievement in these two subjects is also examined. Data from both TIMSS 2019 and Saudi Arabia's National Assessment of Learning Outcomes (NALO) 2018 were employed. Both TIMSS and NALO assess grades 4 and 8 students' mathematics and science achievement, as well as collect a wide range of contextual information about students, their families, teachers, and schools.

This paper contributes to the current literature in two main ways. First, it addresses a need for more evidence about the factors associated with the achievement gap between boys and girls in the MENA region, and especially in Saudi Arabia. The paper explores the surprisingly large and consistent performance gap observed between boys and girls in the region. As shown in figure 1, boys in the region tend to achieve noticeably lower scores than girls in reading, mathematics, and science, with this difference being larger than those in most of the other countries participating in PISA 2018. A further feature of the paper is its examination of two large-scale assessments that have a shared focus on the same domains of study and the same grade levels in a complementary fashion, with data from NALO used to complement findings arising from the analysis of TIMSS data. This paper, therefore, provides important insights for policy makers, both in Saudi Arabia and in other countries in the region, regarding the factors that may contribute to the observed gender gaps in achievement across a range of subjects.

Secondly, Saudi Arabia offers a unique setting in which boys and girls attend separate schools on a universal basis starting from grade 1, being educated only by male and female teachers, respectively, in effect inhabiting parallel education systems, as shown in figure 2, which presents the distribution of grade 4 and grade 8 students in single-gender or mixed education among countries participating in TIMSS 2019.[3] Although gender-segregated schools are not uncommon in the MENA region, students do not usually attend single-gender schools until the end of primary education. The unique structure of the Saudi education system provides an opportunity to examine, in a multilevel framework, how variance in system-level factors applying only to boys or to girls contributes to the observed individual differences in achievement. This analysis exploits the existence of parallel gender-segregated school environments that operate within a shared overarching cultural context, where expectations and practices outside school also vary significantly between boys and girls. While this paper exploits this feature of the Saudi education system in its analysis, it also acknowledges potential difficulties in interpreting findings due to this extreme degree of separation as teacher and school characteristics are confounded with gender differences in learning outcomes. For example, any differences between girls' and boys' educational environments seen in these data are inseparable from the fact that the (male) teachers of boys have been trained and work

[3] In 2019, Saudi Arabia announced that boys would start to be educated by female teachers in grades 1 through 3. These boys' classes are kept separated from the girls' classes. Currently, there are few girls' schools offering boys' classes with female teachers, though the intention is to increase the number.

in an environment that is completely separate to the training and work environment of the (female) teachers of girls.

This paper is organized as follows. Section 2 describes the education system in Saudi Arabia; section 3 discusses the data and methodology; section 4 presents the results from the models; and section 5 includes a discussion of the main findings and their policy implications.

2. Education System in Saudi Arabia

Preuniversity education in Saudi Arabia is divided into four levels: preprimary, elementary, intermediate, and secondary education. Preprimary education includes three years starting at age 3; elementary education starts normally at age 6 and includes grades 1 through 6; intermediate education comprises grades 7 to 9; and secondary education consists of grades 10 through 12. Students in Saudi Arabia attend single-gender schools, except in preprimary. Boys and girls are separated from grade 1 and taught by teachers of the same gender. A recent reform, though still on a limited scale, has allowed boys in grades 1 through 3 to enroll in girls' elementary schools, and, as a result, boys are taught by female teachers but in separate classes.

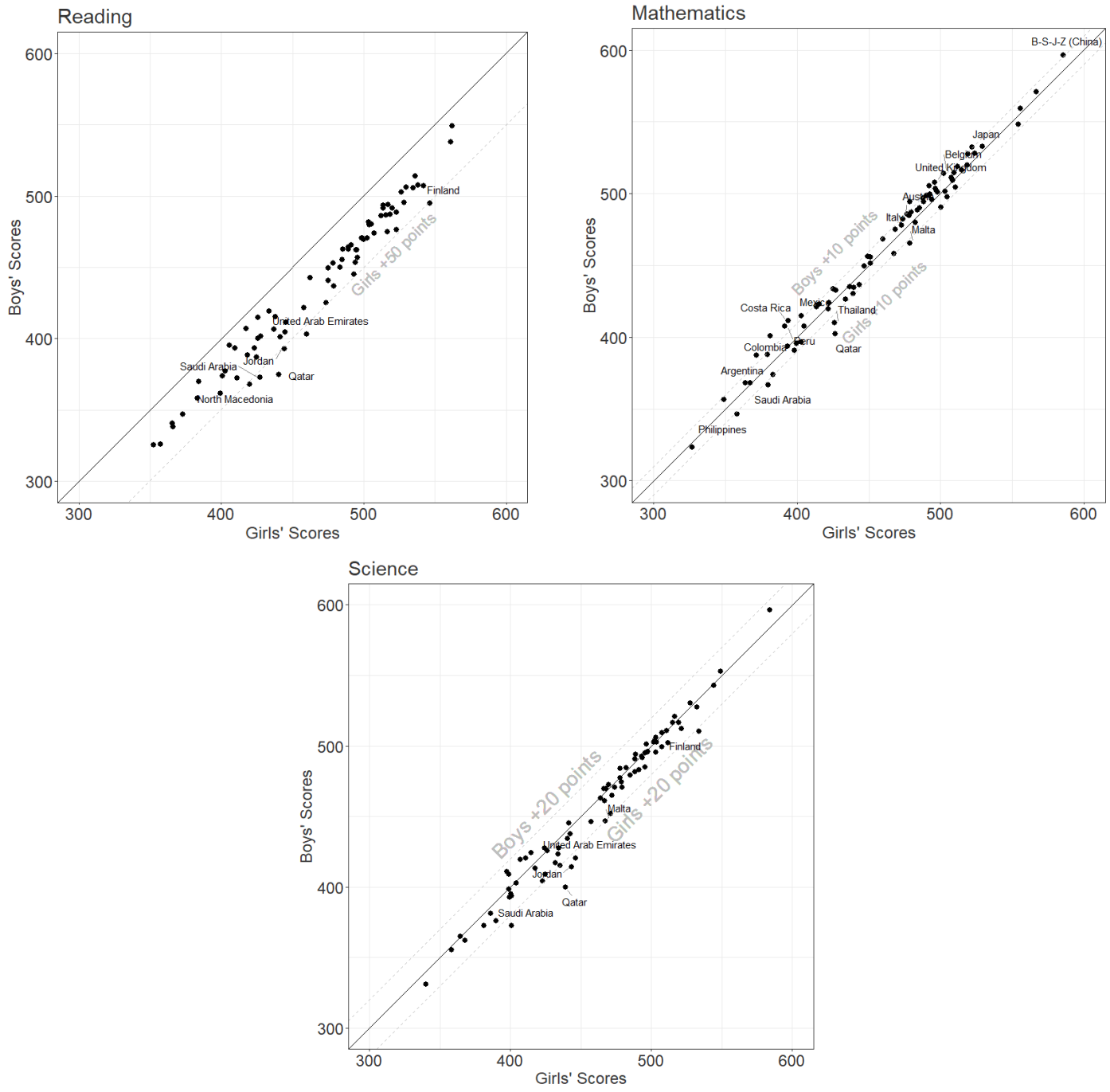
Saudi Arabia's K–12 education system includes more than 5.5 million students and more than 450,000 teachers and is administered through 47 education directorates and 383 education offices within directorates. The Saudi Ministry of Education (MOE) plays a central role in setting the policies and regulations for schools across the country including curriculum, teacher hiring and promotion, and student assessment. Directorates and offices are responsible for implementing the directives of the MOE and tend to have a similar structure across the country (OECD 2020).

Over the last few decades, Saudi Arabia has achieved substantial progress in improving access to education. For example, the gross enrollment ratio (GER) in primary education increased from 58 percent in 1979 to 101 percent in 2019. During the same period, the GER in secondary education increased from 27 percent to 112 percent. Although increased access to education is a positive development, these large gains in access have not been accompanied by similar improvements in students' learning outcomes.

Overall, learning outcomes remain below expectations in Saudi Arabia. Data from TIMSS 2019 show that, in mathematics, Saudi Arabia ranks 53rd of 58 countries (i.e., below 52 countries) in grade 4 and 37th of 39 (outperformed by 36 countries) in grade 8. In PISA 2018, less than half (48 percent) of 15-year-old students in Saudi Arabia achieved minimum proficiency in reading and almost no student was a top performer (i.e., achieving levels 5 or 6; OECD 2019). Additionally, only 27 percent of Saudi students in the same age group achieved at least minimum proficiency in mathematics, compared to an Organisation for Economic Co-operation and Development (OECD) average of 75 percent. In particular, learning outcomes in Saudi Arabia are low relative to the country's level of wealth, and studies (Patrinos and Angrist 2019) have identified Saudi Arabia as an outlier when examining Harmonized Learning Outcomes[4] relative to gross domestic product (GDP) worldwide. Although, as noted above, significant gender differences in achievement can be found in other countries — sometimes in favor of boys, and sometimes in favor of girls (Mullis et al. 2017; 2020) — the gender differences observed in Saudi Arabia are consistently among the largest in the world, with girls showing a consistent advantage, across grade levels and subject areas.

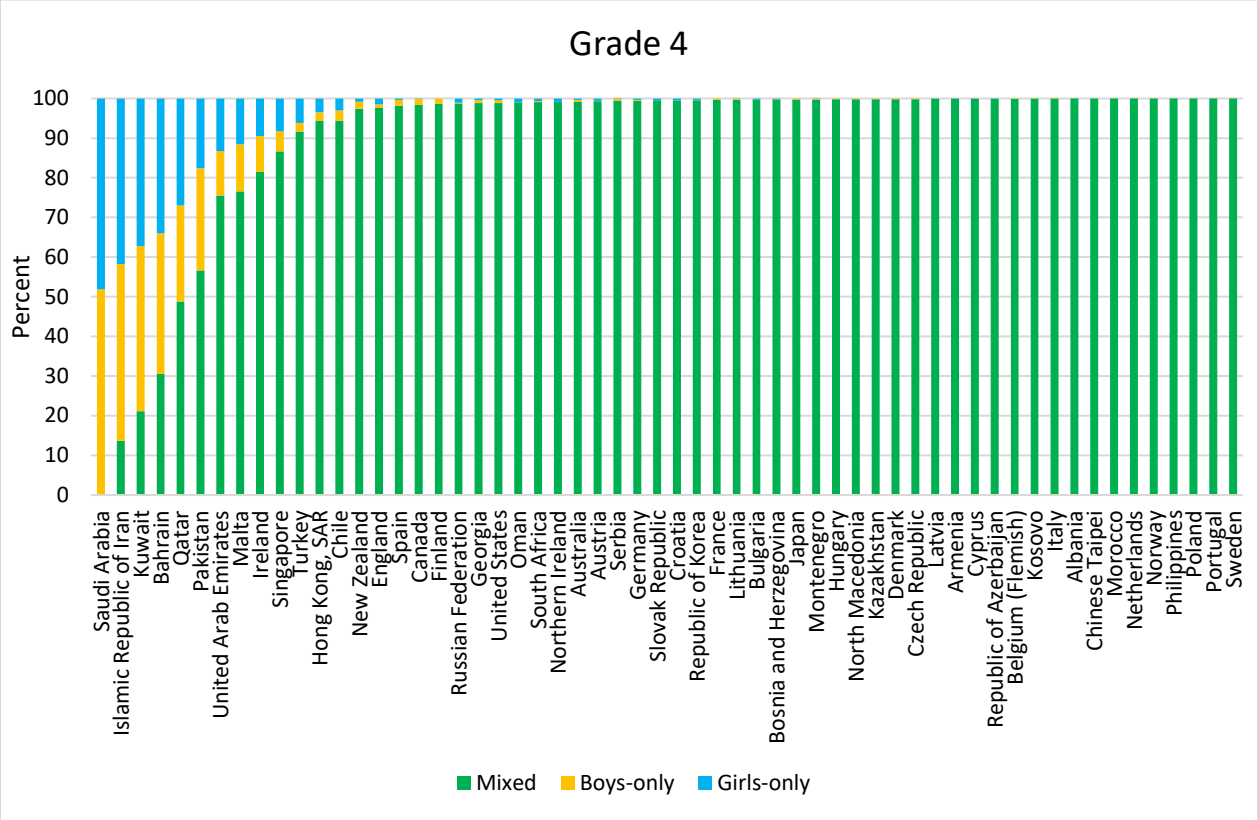
[4] A composite indicator of learning outcomes at the country level, based on data from large-scale assessments such as TIMSS, PISA, Progress in International Reading Literacy Study (PIRLS), and early-grade reading or mathematics assessments (Angrist et al. 2021).

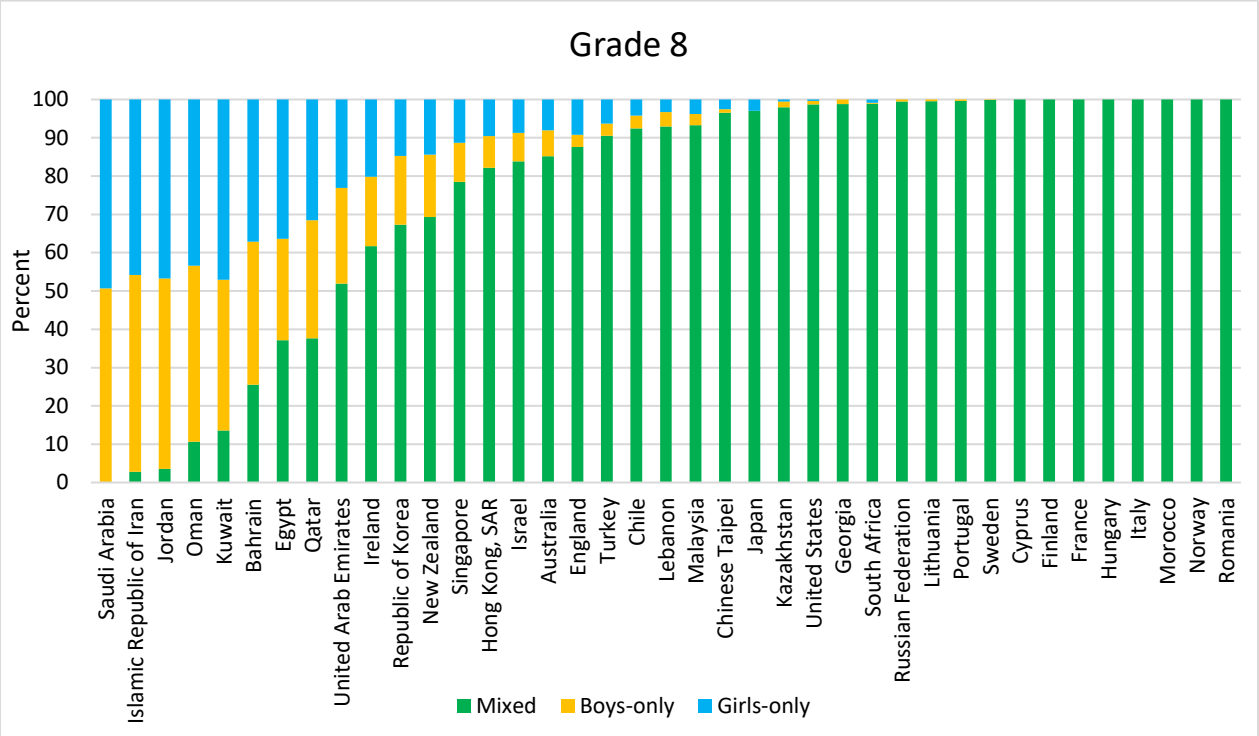
Figure 1. Achievement gap between girls and boys in countries participating in PISA 2018



Source: OECD PISA 2018 data.

Figure 2. Distribution of students in single-gender or mixed education among countries participating in TIMSS 2019





Source: TIMSS 2019 data.

3. Methods

3.1 Data

The analyses described in this paper are drawn from two recent large-scale assessments of education: TIMSS 2019 and Saudi Arabia's NALO 2018. TIMSS provides a robust nationally representative sample, while the NALO study was conducted to provide regionally representative information (requiring a much larger sample size) as well as national-level data. Although the paper focuses on the national level in the analyses described below, it should be noted that subsequent analyses at the regional level would be possible using NALO data.[5] In addition, there is a substantial degree of overlap between the content covered by the TIMSS and NALO contextual questionnaires, although some variables appear in one study but not the other, or they are presented in slightly different formats. The primary analysis, reported in depth in the next section, is conducted using TIMSS 2019 data. Given the high degree of overlap between the two studies' focus on mathematics and science, at the same two grade levels (grades 4 and 8), data from NALO 2018 are used to supplement this primary analysis by drawing on variables of particular interest to Saudi Arabia that have no equivalents in TIMSS. In other words, NALO 2018 data are used as supplementary information to shed additional light on questions arising from the multilevel analysis of TIMSS data. For this reason, NALO data are described in text, with full tabulation of the NALO analyses presented in appendix A1.

3.1.1 TIMSS 2019

TIMSS is a study of the International Association for the Evaluation of Educational Achievement (IEA). It assesses mathematics and science achievement at two grade levels, grades 4 and 8. TIMSS has been carried out every four years since 1995. In 2019, 64 countries participated in TIMSS, including Saudi Arabia. In addition to providing countries with robust data on mathematics and science achievement, TIMSS collects a wealth of contextual data from students, parents, teachers, and school principals. In Saudi Arabia, 5,453 grade 4 students (mean age 9.9 years; 49.6 percent male) across 220 public and private schools and 5,680 grade 8 students (mean age 13.9 years; 49.2 percent male) across 209 schools took part in TIMSS 2019. The TIMSS 2019 data were collected using a stratified two-stage cluster sample design, with a sample of schools selected randomly at the first stage and one or more classes of students selected per each of the sampled schools at the second stage (LaRoche, Joncas, and Foy 2020). The Saudi sample of schools was drawn systematically in order for the sampled schools to represent the populations of grade 4 and grade 8 students nationally, with representation from 13 regions and a balance between male and female schools. Implicit stratification methods were used also to ensure representation of the various school types (such as public versus private schools). The IEA requires high participation rates and adherence to standardized administration procedures for participating countries to be included in the international results. The IEA calculates and provides sampling weights (to ensure that the final sample of participating students can be generalized to the national populations of grade 4 and grade 8 students) and plausible values[6] for mathematics and science scores (to ensure accurate population-level estimates of achievement) in order to facilitate appropriate analyses, taking the complex nature of the data into account.

[5] Representative regional-level data from NALO could be exploited to examine the varying availability of resources and variability in practices across the different regions of Saudi Arabia, and how region-level differences are related to differences in achievement and the gender gap.

[6] Plausible values are generated by imputing a set of values (five values in the case of TIMSS) representing 'plausible' estimates of student achievement based on their responses to the assessment and background variables. Plausible values are not suitable for reporting individual-level results, but at the population level the use of plausible values facilitates the calculation of appropriate standard errors for complex survey designs such as those used by TIMSS where each student is administered only a small subset of the items in the assessment.

3.1.2 NALO 2018

Saudi Arabia's NALO is administered by the Education and Training Evaluation Commission (ETEC), an independent government agency responsible for school evaluation, accreditation, and assessment, among other responsibilities. In 2018, the domains assessed by NALO were mathematics and science in grades 4 and 8, which means that the data from NALO 2018 are closely aligned to TIMSS 2019 both in terms of the target domains and grade levels. In NALO 2018, 27,985 grade 4 students (50.2 percent male) across 964 government, private, and Quran schools completed tests of mathematics and science, as did 30,157 grade 8 students (49.6 percent male) across 939 schools. The schools that took part in NALO were sampled using procedures similar to those used in TIMSS.

3.2 Statistical analysis

Prior to the main analysis, descriptive statistics were computed, and relevant statistical tests were conducted to provide a comprehensive overview of the gender differences across the contextual variables of interest to this study. The levels of statistical significance along with the relevant effect sizes for each of these differences are reported. The phi (φ) and the Cramer's V (φ_c) effect size measures were used for the contextual categorical variables for 2x2 contingency tables and for contingency tables larger than 2x2, respectively. The Cohen's *d* effect size measure was used for the contextual continuous variables (Fritz, Morris, and Richler 2012). Cohen (1988) guidelines were used in conjunction with Hattie (2009) guidelines for the interpretation of the effect sizes. The IEA International Database Analyzer (IDB Analyzer) and the Statistical Package for the Social Sciences (SPSS) were used to compute descriptive statistics for TIMSS and NALO, respectively.

Next, multilevel linear regression analysis was used to explore the different factors that are associated with boys' underperformance in Saudi Arabia as well as all students' achievement in mathematics and science in grades 4 and 8 more generally. The multilevel linear regression models draw on the TIMSS 2019 data, with NALO 2018 data used to supplement the analysis by describing data that are not available in TIMSS. Four separate families of models were constructed: grade 4 mathematics, grade 4 science, grade 8 mathematics, and grade 8 science. All plausible values of achievement and sampling weights were used in all the analyses as per the relevant guidelines by von Davier et al. (2009) and Rutkowski et al. (2010), respectively. Assumptions necessary for conducting the multilevel linear regression analysis (i.e., linearity, homogeneity of variance, normality of errors) were checked and met, and parameters for the models were estimated using the maximum likelihood estimation with robust standard errors (*SEs*) (Muthén and Muthén 2017). Multilevel linear regression analysis was performed using Mplus 8 software.

Along with mathematics and science achievement scores, variables included in the multilevel models are drawn from the student and home questionnaires (level 1 of the analysis) as well as the teacher and school principal questionnaires (level 2 of the analysis). Variables were selected for inclusion in the models *a priori*, based on the previous literature on student achievement and the gender gap in education in particular, as discussed in section 1 of this paper, in addition to the expected theoretical or policy relevance of these variables to the question of gender differences in performance in the Saudi context. As far as possible, each model was constructed using the same set of variables, notwithstanding some slight differences arising from the selection of variables related specifically to mathematics or science instruction and some differences between the grade 4 and 8 questionnaires.

A hierarchical approach was followed, whereby conceptually similar variables were entered into each step of each model in blocks, as shown in table 1. The first step of each model included student gender only. By including the student gender variable into the model alone, the difference in achievement between boys and girls, after controlling for the clustering of the data within classes/schools, could be observed. Next, different blocks of variables were entered into the model one by one to allow both for the examination of their contribution in explaining the difference in achievement between boys and girls

as well as their contribution in predicting overall achievement. In the final step of each model, the statistical significance of the interactions between student gender and each of the predictor variables in predicting achievement was explored. Each interaction term was entered into the model individually and all the statistically significant interaction terms were entered into the final model. To facilitate interpretation of the statistically significant interaction terms, those were plotted using the predicted values based on the last step (step 7) of each model. Hence, the interaction plots in the next section present the predicted, rather than the raw, gender differences in each variable in terms of mathematics and science achievement, after accounting for a range of student- and class/school-level predictor variables.

Table 1. Steps in building the hierarchical two-level linear regression models

	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7
Student gender	✓	✓	✓	✓	✓	✓	✓
Demographics and home background		✓	✓	✓	✓	✓	✓
Student engagement and attitudes			✓	✓	✓	✓	✓
School climate				✓	✓	✓	✓
Teacher qualifications and practices					✓	✓	✓
School leadership and resources						✓	✓
Interaction terms							✓

Note: ✓ indicates variables included in each step.

Equation 1 represents the null multilevel models (i.e., models with no predictor variables), which were applied to the TIMSS data, controlling for their clustering and allowing for the estimation of the proportions of the total variance in the outcome variable that is attributable within and between clusters (i.e., intraclass correlation (*ICC*) coefficients).

$$y_{ij} = \beta_0 + u_{0j} + e_{ij} \quad (1)$$

where y_{ij} is the outcome variable (e.g., mathematics achievement) of student *i* in class/school *j*, β_0 is the mean intercept, u_{0j} is the variation of class/school *j* from the mean intercept, and e_{ij} is the student-level residual error term.

Equation 2 represents the multilevel models, which were applied to the TIMSS data, including v number of predictor variables while controlling for the clustering of the data.

$$y_{ij} = \beta_0 + \beta_{1j}x_{1ij} + \ldots + \beta_{vj}x_{vij} + u_{0j} + e_{ij} \quad (2)$$

where y_{ij} is the outcome variable (e.g., mathematics achievement) of student i in class/school j, β_0 is the mean intercept, β_{1j} is the regression slope for the predictor variable x_1 of student i in class/school j, β_{vj} is the regression slope for the predictor variable x_v of student i in class/school j, u_{0j} is the variation of class/school j from the mean intercept, and e_{ij} is the student-level residual error term.

Reported statistics for each of the models include: proportions of variance (R^2; expressed as a percentage of the total variance) in achievement explained at each level and step; intercepts with their *SEs*; unstandardized coefficients (*Bs*) and standardized coefficients (*ßs*) each accompanied by their *SEs* for each predictor variable; fit statistics (Loglikelihood *(H_0)*, Akaike Information Criterion *(AIC)*, and Bayesian Information Criterion *(BIC)*) for each step, including the null model; and *ICC* coefficients for the null model. *Bs* are expressed in the original unit of each of the predictor variables, while *ßs* can be used to compare

the relative strength of each predictor variable in predicting achievement (i.e., to find the most robust predictors of achievement) in each model. Also, although fit indices are not intrinsically interpretable (i.e., their values cannot be interpreted as being large or small in themselves), they can be compared across different steps of each model to check whether changes in the model lead to better fit; for all three indices presented in the tables, smaller values indicate better model fit regardless of the absolute number.

4. Results

4.1 Descriptive statistics

4.1.1 Descriptive statistics of outcome variables

Table 2 shows the mean scores in mathematics and science by student gender in the TIMSS 2019 and NALO 2018 assessments. Results from both data sets show that boys consistently underperform compared to girls in both subject domains and at both grade levels in Saudi Arabia. The differences are larger for science than for mathematics at both grade levels, in both TIMSS and NALO.

Table 2. Mean mathematics and science achievement in Saudi Arabia, by grade level and student gender, TIMSS 2019 and NALO 2018

		TIMSS 2019			NALO 2018		
		Male *Mean (SD)*	Female *Mean (SD)*	*M-F*	Male *Mean (SD)*	Female *Mean (SD)*	*M-F*
Grade 4	Mathematics	385 (108)	412 (91)	*-26*	496 (103)	504 (97)	*-8*
	Science	373 (116)	434 (97)	*-60*	483 (101)	518 (96)	*-35*
Grade 8	Mathematics	385 (80)	403 (74)	*-17*	493 (101)	507 (99)	*-14*
	Science	408 (91)	455 (79)	*-47*	480 (100)	521 (96)	*-42*

Source: TIMSS 2019 data reported by Mullis et al. (2020). NALO data are authors' own calculations. All differences between genders have been rounded.

4.1.2 Descriptive statistics of predictor variables (TIMSS 2019)

Table 3 shows the percentages of boys and girls in grades 4 and 8 in Saudi Arabia across the TIMSS 2019 contextual categorical variables of interest to this paper (i.e., categorical variables that were included in the models). Table 4 shows the means (*M*), standard deviations (*SD*), minimum (*min*), and maximum scores (*max)* for boys and girls in grades 4 and 8 in Saudi Arabia across the TIMSS contextual continuous variables of interest to this study (i.e., continuous variables that were included in the models). The tables also include the effect sizes (φ/φ_c and *d*) for the gender differences in each of the contextual variables (see section 3.2 for further information). Given that some of the estimates presented in table 4 are not intrinsically interpretable (i.e., their values cannot be interpreted as being large or small per se), cut-scores and their corresponding interpretation along the continuum of each of these variables as set by TIMSS are provided in table A2.1 in appendix A2.

Across most of the TIMSS 2019 contextual variables, the differences between boys and girls are statistically significant; however, most of the differences yielded small to moderate effect sizes. Among these differences the most noticeable at grade 4 are observed in teachers' absenteeism, with girls being more likely to attend schools where teacher absenteeism is a minor problem compared to boys who are more likely to attend schools where teacher absenteeism is either not a problem or a moderate to serious problem. Another considerable difference can be found in teachers' major area of study, with more teachers in boys' schools having education and mathematics or science as their major area of study, compared to teachers in girls' schools who tend to primarily have mathematics or science but not education as their major area of study. Differences in teachers' professional development are also noticeable, with teachers in boys' schools reporting having attended fewer hours of professional development on mathematics and science compared to teachers in girls' schools. Also, in comparison to grade 4 girls, grade 4 boys are less likely to have positive attitudes (liking/confidence) toward mathematics

and science, and a lower sense of school belonging, plus they experience more frequent bullying. Additionally, boys' schools tend to be less safe and orderly, and teachers in boys' schools tend to also report lower levels of job satisfaction. At grade 8, both mathematics and science teachers are younger, on average, in boys' schools compared to teachers in girls' schools, while, in line with findings at grade 4, more grade 8 teachers in boys' schools have education and mathematics or science as their major area of study, compared to teachers in girls' schools, who tend to primarily have mathematics or science but not education as their major area of study. Grade 8 teachers in boys' schools also report having attended fewer hours of professional development on mathematics and science compared to teachers in girls' schools. Additionally, grade 8 boys' schools tend to be less safe and orderly, with lower levels of discipline and more frequent bullying among students compared to girls' schools, while teachers in boys' schools tend to also report lower levels of job satisfaction.[7]

Table 3. Contextual categorical variables by student gender, TIMSS 2019

	Grade 4			Grade 8		
	Male	Female	φ/φ_c	Male	Female	φ/φ_c
Student-level variables	%	%		%	%	
Student immigration status			0.029			0.051**
native	64.8	66.9		76.7	81.4	
second-generation immigrant students	28.8	27.6		15.4	11.7	
first-generation immigrant students	6.4	5.5		7.9	6.8	
Student owns mobile phone			0.036**			0.000
yes	62.7	58.3		81.4	82.3	
no	37.3	41.7		18.6	17.7	
Preschool attendance and duration			0.018			
did not attend	30.8	30.0				
1 year or less	39.7	39.0				
2 years	18.3	18.3				
3 years or more	11.2	12.7				
Student absenteeism			0.131***			0.182***
never or almost never	36.6	47.8		34.3	20.0	
once every two months	10.3	9.9		15.8	11.7	
once a month	12.3	10.5		17.7	21.7	
once every two weeks	10.0	7.3		13.7	21.5	
once a week	30.8	24.5		18.6	25.1	
Time spent on mathematics homework						0.051**
15 minutes or less				75.8	73.3	
16 minutes or more				24.2	26.7	
Time spent on science homework						0.045**
15 minutes or less				71.0	77.4	
16 minutes or more				29.0	22.6	

[7] The TIMSS variable describing teachers' qualifications for grade 4 shows that a majority of teachers in Saudi Arabia have secondary education. This is not consistent with official data from the Ministry of Education or other available data sources, which indicate that most teachers in the country have at least a bachelor's degree. Due to this inconsistency, the authors of the study decided to omit this variable from the analysis.

	Grade 4			Grade 8		
	Male	Female	φ/φ_c	Male	Female	φ/φ_c
Class/school-level variables	%	%		%	%	
School location			0.068***			0.247***
urban	61.5	58.7		66.4	47.9	
suburban/medium size city or large town	27.5	27.4		15.0	33.7	
small town or village/remote rural	11.0	13.9		18.6	18.4	
Teacher age (mathematics teacher)			0.073***			0.243***
29 years or younger	9.7	4.8		28.3	3.9	
30–39 years	39.6	47.0		48.2	62.1	
40 years or older	50.8	48.2		23.5	34.0	
Teacher age (science teacher)			0.047**			0.126***
29 years or younger	12.7	6.5		10.2	2.1	
30–39 years	38.9	46.8		56.2	50.7	
40 years or older	48.4	46.7		33.6	47.2	
Time assigned to mathematics homework			0.147***			0.119***
15 minutes or less	52.7	67.2		58.4	62.8	
16 minutes or more	47.3	32.8		41.6	37.2	
Time assigned to science homework			0.054**			0.180***
15 minutes or less	77.3	69.8		66.8	81.4	
16 minutes or more	22.7	30.2		33.2	18.6	
Poor teacher timekeeping			0.157***			0.165***
not a problem	42.0	52.1		39.3	56.6	
minor problem	35.6	34.0		42.2	26.5	
moderate or serious problem	22.4	13.9		18.5	16.9	
Teacher absenteeism			0.211***			0.066***
not a problem	45.0	33.1		47.8	42.8	
minor problem	28.6	48.9		31.8	32.0	
moderate or serious problem	26.4	18.0		20.4	25.2	
Teacher highest level of education (mathematics teacher)						0.174***
Up to ISCED level 6 — bachelor's or equivalent level (grade 8)				97.7	100.0	
ISCED levels 7 & 8 — master's or doctorate degree (grade 8)				2.3	0.0	
Teacher highest level of education (science teacher)						0.038**
Up to ISCED level 6 — bachelor's or equivalent level (grade 8)				91.3	98.0	
ISCED levels 7 & 8 — master's or doctorate degree (grade 8)				8.7	2.0	
Teacher major area of study (mathematics teacher)			0.222***			0.209***
education and mathematics (grade 4)	40.7	21.0				
mathematics but not education (grade 4)	47.4	59.6				
mathematics and mathematics education (grade 8)				24.9	11.6	
mathematics but not mathematics education (grade 8)				68.3	70.6	
all other majors	11.9	19.4		6.8	17.8	
Teacher major area of study (science teacher)			0.245***			0.167***
education and science (grade 4)	40.9	15.8				
science but not education (grade 4)	49.1	66.9				
science and science education (grade 8)				27.2	10.3	
science but not science education (grade 8)				65.6	85.0	
all other majors	10.0	17.3		7.2	4.7	

	Grade 4			Grade 8		
	Male	Female	φ/φ_c	Male	Female	φ/φ_c
Class/school-level variables	%	%		%	%	
Teacher major area of study (science teacher)			0.245***			0.167***
education and science (grade 4)	40.9	15.8				
science but not education (grade 4)	49.1	66.9				
science and science education (grade 8)				27.2	10.3	
science but not science education (grade 8)				65.6	85.0	
all other majors	10.0	17.3		7.2	4.7	
Professional development hours on mathematics			0.326***			0.196***
none	10.5	4.6		11.9	3.8	
less than 6 hours	20.1	5.4		14.2	12.2	
6–15 hours	34.0	28.0		34.4	26.1	
16–35 hours	19.2	27.1		20.9	31.3	
more than 35 hours	16.2	34.9		18.5	26.7	
Professional development hours on science			0.238***			0.324***
none	18.6	8.7		13.5	0.2	
less than 6 hours	12.4	13.0		12.8	8.9	
6–15 hours	32.9	28.1		35.2	26.0	
16–35 hours	21.1	16.9		16.9	34.4	
more than 35 hours	15.0	33.3		21.5	30.6	
Professional development on mathematics content			0.007			0.098***
yes	55.3	55.0		42.0	50.0	
no	44.7	45.0		58.0	50.0	
Professional development on science content			0.131***			0.182***
yes	43.5	56.1		43.6	62.3	
no	56.5	43.9		56.4	37.7	
Professional development on mathematics pedagogy			0.075***			0.081***
yes	56.8	61.8		60.5	67.1	
no	43.2	38.2		39.5	32.9	
Professional development on science pedagogy			0.053**			0.172***
yes	46.3	47.1		57.2	76.8	
no	53.7	52.9		42.8	23.2	
Principal highest level of education			0.112***			0.027*
ISCED Level 6 — bachelor's or equivalent level	82.4	92.9		90.9	93.7	
ISCED levels 7 & 8 — master's or doctorate degree	17.6	7.1		9.1	6.3	
Principal qualification in educational leadership			0.168***			0.233***
yes	25.6	14.8		28.5	8.5	
no	74.4	85.2		71.5	91.5	
School library			0.187***			0.026
yes	69.5	49.4		74.7	72.4	
no	30.5	50.6		25.3	27.6	

Note: *$p < .05$; **$p < .01$; ***$p < .001$.

Table 4. Contextual continuous variables by student gender, TIMSS 2019

	Grade 4									Grade 8								
	Male				Female				*d*	Male				Female				*d*
Student-level variables	*m*	*SD*	*min*	*max*	*m*	*SD*	*min*	*max*		*m*	*SD*	*min*	*max*	*m*	*SD*	*min*	*max*	
Home resources for learning (grade 4)	9.3	1.41	3.8	14.9	9.5	1.34	3.8	14.9	0.109									
Home educational resources (grade 8)										9.4	1.59	4.6	13.5	9.5	1.51	4.6	13.5	0.058
Literacy and numeracy readiness for school	10.1	2.01	3.1	14.6	10.5	1.91	3.1	14.6	0.235***									
Student likes learning mathematics	10.3	2.03	3.9	13.1	11.3	1.88	3.9	13.1	0.501***	10.3	1.97	5.1	13.9	10.0	2.08	5.1	13.9	0.143**
Student likes learning science	10.0	2.33	2.7	13.2	11.3	2.15	2.7	13.2	0.575***	10.4	2.09	3.9	13.5	10.7	2.15	3.9	13.5	0.146*
Student confident in mathematics	10.1	2.08	2.8	14.4	11.1	2.17	2.8	14.4	0.470***	10.5	1.88	3.3	15.9	10.5	2.01	3.3	15.9	0.010
Student confident in science	10.0	2.01	3.4	13.3	11.0	2.05	3.4	13.3	0.517***	10.5	1.92	3.3	14.8	10.9	2.02	3.5	14.8	0.223***
Student sense of school belonging	10.0	2.33	3.1	12.8	11.1	2.01	3.1	12.8	0.533***	10.2	2.09	3.9	13.3	10.3	1.92	3.9	13.3	0.075
Bullying	8.9	2.25	2.9	12.7	10.2	2.14	2.9	12.7	0.619***	9.8	2.35	2.0	12.9	10.6	1.97	3.6	12.9	0.360***
Class/school-level variables																		
School mean of home resources for learning (grade 4)	9.2	0.83	6.3	11.3	9.5	0.68	7.2	11.3	0.409**									
School mean of home educational resources (grade 8)										9.4	0.69	7.9	11.4	9.5	0.67	8.3	11.6	0.132
Safe and orderly schools (mathematics teacher)	10.8	1.83	5.8	13.4	11.9	1.71	7.5	13.4	0.644***	11.2	1.96	7.2	13.9	11.9	1.92	6.3	13.9	0.376*
Safe and orderly schools (science teacher)	10.9	2.10	5.1	13.4	11.7	1.95	3.9	13.4	0.390*	10.9	1.88	6.3	13.9	11.7	2.11	4.9	13.9	0.410**
School emphasis on academic success	10.8	2.12	4.2	16.4	11.8	2.14	7.2	16.4	0.446**	10.8	1.89	6.7	16.4	11.5	2.03	5.8	16.4	0.337*
School discipline	9.5	2.40	3.7	12.8	10.6	2.15	3.7	12.8	0.443*	10.0	2.68	4.1	14.0	11.6	2.37	4.3	14.0	0.616**
Teacher years of experience (mathematics teacher)	15.9	8.43	0.0	31.0	14.4	7.70	0.0	33.0	0.183	10.6	7.29	0.0	30.0	12.2	7.27	1.0	30.0	0.216
Teacher years of experience (science teacher)	15.3	9.36	0.0	36.0	12.8	7.86	0.0	31.0	0.295	13.1	7.43	0.0	36.0	12.6	7.53	0.0	30.0	0.066
Teacher job satisfaction (mathematics teacher)	10.6	1.36	6.4	11.7	11.2	0.87	8.3	11.7	0.569***	10.7	1.50	5.3	11.8	11.3	0.74	7.2	11.8	0.465**
Teacher job satisfaction (science teacher)	10.6	1.50	5.8	11.7	11.3	0.88	6.9	11.7	0.553***	10.5	1.65	5.6	11.8	11.5	0.64	8.5	11.8	0.783***
Principal years of experience	9.9	6.65	0.0	30.0	8.6	7.09	0.0	33.0	0.185	9.4	7.56	0.0	32.0	8.6	8.30	0.0	39.0	0.102

Note: *$p < .05$; **$p < .01$; ***$p < .001$.

4.1.3 Additional contextual information for grades 4 and 8 (NALO 2018)

All of the differences between boys and girls in grades 4 and 8 noted in this section are statistically significant at the $p < .001$ level. However, the effect sizes associated with these differences were generally small (appendix A1). NALO data indicate that 14.5 percent of grade 4 boys had repeated a year at school because of poor academic performance, compared to 6.6 percent of grade 4 girls (table A1.1). In addition, grade 4 boys reported engaging in a lower level of reading than grade 4 girls. For example, 22.6 percent of boys reported never having read a book, compared to 12.0 percent of girls. Conversely, 33.8 percent of boys have read more than 10 books, compared to 40.4 percent of girls. However, when asked whether mathematics is important in life, grade 4 boys and girls provided broadly similar levels of agreement (table A1.1). Among grade 4 immigrant students (those reported by their parents to have been born outside Saudi Arabia), boys were more likely to attend a private school (13.9 percent) than girls (2.6 percent) (table A1.2). However, the vast majority of both boys (85.0 percent) and girls (95.2 percent) attended government schools.

NALO data also show the differences in perceptions among teachers, school principals, and parents in boys' and girls' schools. Teachers of grade 4 boys are substantially less likely to agree that parents have a good understanding of their child's current academic level, suggesting greater misalignment between student performance and parental understanding in boys' schools compared to girls' schools. This difference was associated with the largest effect size observed among all the selected NALO variables ($\varphi = .302$; table A1.3). Teachers in boys' schools are also less likely to report that high-achieving students were respected among their peers compared to teachers in girls' schools (table A1.3). School principals' reports correspond with those of their teachers in relation to parental support for learning. Principals of girls' schools report a higher degree of parental support for learning than those in boys' schools, and also a higher degree of satisfaction among parents with their child's educational progress (table A1.4). More grade 4 boys (67.3 percent) have access to a school library from which they can borrow compared to grade 4 girls (48.5 percent), despite girls reporting reading more books, as noted above.

Similar to grade 4, boys in grade 8 are more likely (6.9 percent) than girls (4.5 percent) to report having repeated a year at school because of poor academic performance (table A1.1). However, both in absolute and in relative terms, the differences are smaller at grade 8 than grade 4. In terms of reading behavior, grade 8 boys report a more nuanced pattern than seen at grade 4. As at the lower grade, more boys (30.6 percent) than girls (19.3 percent) report never having read a book. However, boys and girls are equally likely to report having read more than six or more than 10 books. Another difference from students' responses at grade 4 is that, in grade 8, boys agree more strongly that mathematics is important in life relative to their female peers (47.6 percent of boys agreeing a lot compared to 36.3 percent of girls). Among immigrant students, grade 8 boys were much more likely to attend a private school (11.1 percent) than grade 8 girls (2.7 percent) (table A1.2). Nonetheless, as at grade 4, most boys (87.7 percent) and girls (96.3 percent) attended government schools.

Again, similar to grade 4, NALO data for grade 8 show the differences in perceptions among teachers, school principals, and parents in boys' and girls' schools. A substantially lower proportion of teachers of grade 8 boys agree that parents have a good understanding of their child's current academic level (associated with the second-largest effect size observed: $\varphi = .236$; table A1.3). Similar to grade 4, this suggests greater misalignment between student performance and parental understanding among parents in boys' schools. A lower level of respect for students who achieve at a high academic level is reported in boys' schools, although teachers in boys' schools are slightly more likely to view their grade 8 students as always being keen to excel academically. School principals report weaker parental support for learning for grade 8 boys than for grade 8 girls (table A1.4), and that boys' parents' expectations are being met to a lesser extent. There is little difference in grade 8 boys' and girls' access to a school library from which they can borrow books.

4.2 Overall findings from multilevel models

In this section, the main findings on boys' underperformance in Saudi Arabia and factors associated with student achievement, in general, are first presented. Then, results from the multilevel models are shown in detail for grades 4 and 8 separately. Table 5 summarizes the main findings from the analysis. As discussed in the methods section, a series of hierarchical two-level linear regression models were run, starting with a simple model that includes student gender but no other predictor variables. Then, with each step, the changes in the achievement gap between girls and boys are explored when adding additional information on student demographics and home background, student engagement and attitudes, school climate, teacher qualifications, and school leadership and resources. Table 5 shows the coefficient on the gender gap. Regression estimates for the other predictor variables are presented in appendix A3.

As shown in table 5, boys underperform compared to girls in Saudi Arabia across both grades and subjects (step 1). The achievement gap between boys and girls is greater in science than in mathematics across both grades. For example, boys in grade 4 underperform girls by 53 points in science compared to 20 points in mathematics. The results also show that, in grade 4, controlling for student, teacher, and school characteristics accounts for the entire gender gap in mathematics and more than half of the gap in science. Specifically, when student-level predictor variables such as student demographics, home resources for learning, and literacy and numeracy readiness[8] are taken into account, the gap in mathematics between girls and boys drops from 20 points to 8 points and it is no longer statistically significant (step 2).

However, in grade 8, controlling for a wide range of observable characteristics from the student, parent, teacher, and principal questionnaires explains a relatively small portion of the achievement gap between boys and girls in both mathematics and science. As shown in table 5, the gender gap between grade 8 boys and girls declines by 4 points in mathematics and 11 points in science once all the predictor variables are included (step 6). However, a significant unexplained gender gap favoring girls still exists in both subjects in grade 8. Estimates from table 5 show that boys underperform girls by 16 points in mathematics and 40 points in science, even after controlling for all observable characteristics. Some potential explanations for this remaining gender gap are discussed in section 5.

Table 5. Summary of main findings of the hierarchical two-level linear regression models

	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
Student gender (female = 1)						
Grade 4 mathematics	19.73* (8.67)	7.68 (7.95)	-7.42 (7.98)	-17.55 (8.99)	-18.68 (12.51)	-6.71 (11.92)
Grade 4 science	53.14 (8.29)***	39.36 (7.04)***	21.61 (6.57)**	17.44 (7.38)*	15.04 (8.57)	21.52 (8.79)*
Grade 8 mathematics	20.78 (5.29)***	15.79 (4.78)**	18.14 (4.52)***	15.47 (4.92)**	9.88 (6.06)	16.29 (6.98)*
Grade 8 science	50.40 (5.80)***	45.83 (5.26)***	42.28 (5.00)***	35.42 (5.03)***	34.88 (7.10)***	39.89 (7.52)***
Demographics and home background		✓	✓	✓	✓	✓
Student engagement and attitudes			✓	✓	✓	✓
School climate				✓	✓	✓
Teacher qualifications and practices					✓	✓
School leadership and resources						✓

Note: ✓ indicates variables included in each step.

[8] 'Literacy and numeracy readiness' is a TIMSS composite scale based on parental reports of the extent to which their child could demonstrate specified literacy and numeracy skills when they started school (e.g., read some words, write their own name, count by themselves, recognize written numbers). The scale is available only for grade 4 students, as the parents of grade 8 students were not asked to complete a questionnaire.

To explore the extent to which predictors in the models have different effects on boys' and girls' performance, the statistical significance of the interactions between student gender and each of the predictor variables is examined in step 7 of each model. The results from these interactions are presented in the final step (step 7) of each model, in tables 6 and 7. The final models explained a substantial proportion of observed achievement variance across both grades and subjects. For example, the final model for grade 4 mathematics explains 71 percent of variance at the class/school level and 26 percent of variance at the student level, or 40 percent of the total observed variance.

Overall, results for the examined interaction terms show that, in grade 4, school climate, student absenteeism, and early numeracy and literacy skills contribute to the achievement gap between girls and boys in Saudi Arabia. A safe and orderly school climate is more strongly associated with improvements in boys' mathematics and science achievement than girls' achievement (in the two subjects). The findings also indicate that boys' mathematics achievement decreases to a greater degree than girls' achievement with more frequent student absenteeism. In addition, the results suggest that, even though greater literacy and numeracy readiness was linked with improvements in science achievement of both boys and girls, boys tended to benefit more from this readiness than girls. For grade 8, boys' mathematics achievement increases to a greater degree in schools with stronger emphasis on academic success than girls' achievement. Higher levels of confidence in science were also associated with greater achievement gains in the subject among boys compared to girls.

4.3 Findings for grade 4

In this section, the results from the final two steps (steps 6 and 7) of each model are presented for grade 4 mathematics and science. Results for all steps are shown in appendix A3.

4.3.1 Mathematics

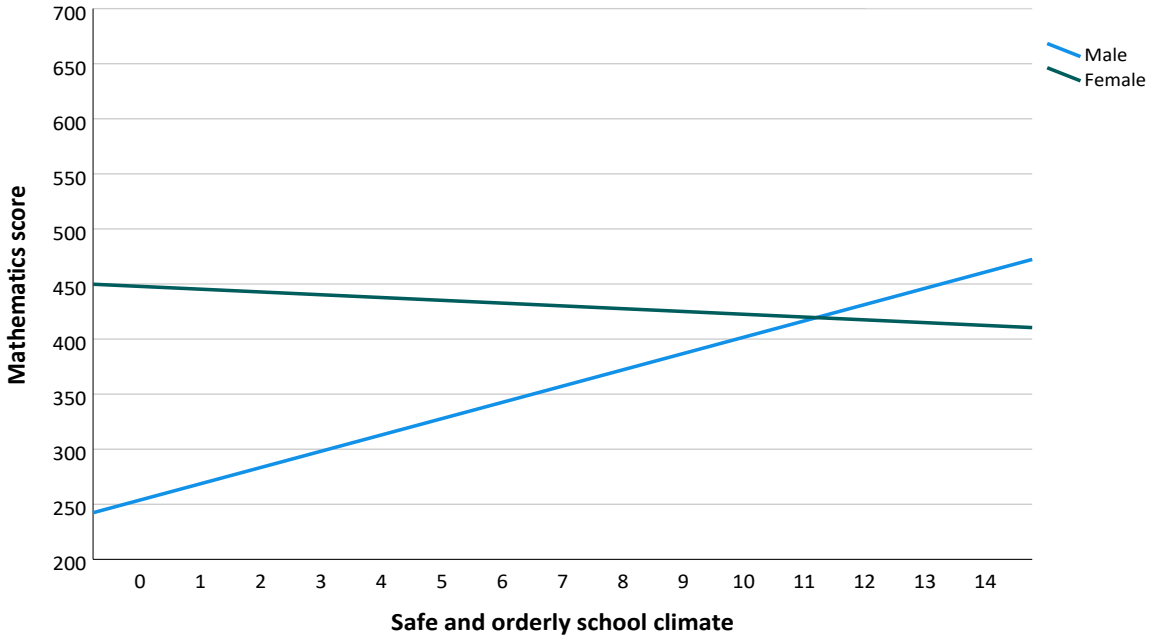
Table 6 provides the coefficients and model statistics for grade 4 mathematics from steps 6 and 7. Step 7, the final model, explains 71 percent of the class/school-level variance and 26 percent of the student-level variance, or 40 percent of the total observed variance. As shown in step 6, the coefficient of student gender is negative and statistically insignificant. This indicates that when controlling for student, home, teacher, and school characteristics, the achievement gap between grade 4 boys and girls in mathematics becomes statistically insignificant. Findings from step 6 also suggest that students' home resources for learning, early literacy and numeracy skills, absenteeism, bullying, attitudes toward mathematics, and sense of school belonging[9] are significantly associated with student achievement. For instance, students who are absent once a week tend to underperform students who are never or almost never absent by 24 points, which is equivalent to 24 percent of a standard deviation. Also, students with stronger early literacy and numeracy skills achieved higher mathematics scores (B = 7.6, p < .001) relative to other students. Students who reported being bullied less frequently achieved higher mathematics scores (B = 4.9, p < .001). At the class/school level, school location, poor teacher timekeeping, teacher experience, and professional development are significantly associated with student performance. Surprisingly, after holding other variables constant, students in schools located in small towns or remote areas perform better in mathematics than students in urban areas (B = 54.9, p < .05), while with each additional year of teacher experience, students score 1.8 points higher in mathematics (p < .01).

In step 7, the interactions between student gender and each of the predictors are examined. However, only the results for interaction terms that are statistically significant are reported. Overall, poor school climate tends to affect boys more negatively than girls. As shown in figure 3, the achievement gap

[9] Surprisingly, the coefficient of sense of school belonging is negative, suggesting a negative correlation between student achievement and sense of school belonging. One potential explanation is that high-achieving students in Saudi Arabia may feel alienated within schools or not appreciated/respected by their peers, which is consistent with some of previous literature discussed in the introduction (e.g., Jha and Pouezevara 2016).

between boys and girls is greater in schools with poor school climate relative to other schools. In schools with a safe and orderly school climate, boys and girls tend to perform similarly.

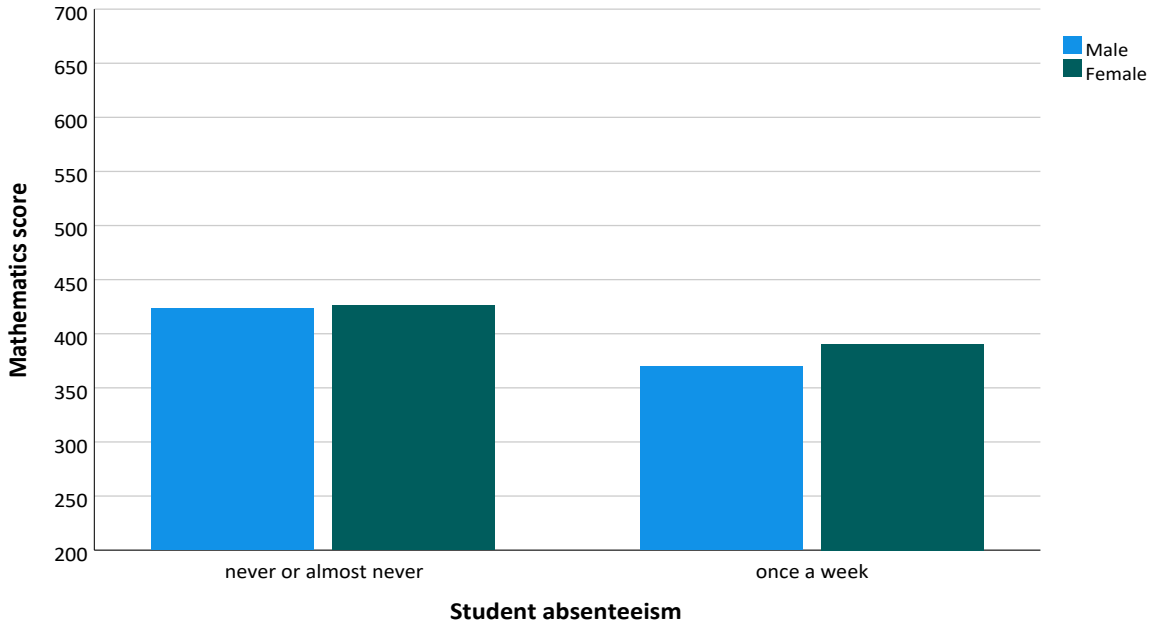
Figure 3. Interaction between student gender and safe and orderly school climate on mathematics achievement, grade 4



Note: The plot presents the predicted values from step 7 of the mathematics model for grade 4 (table 6).

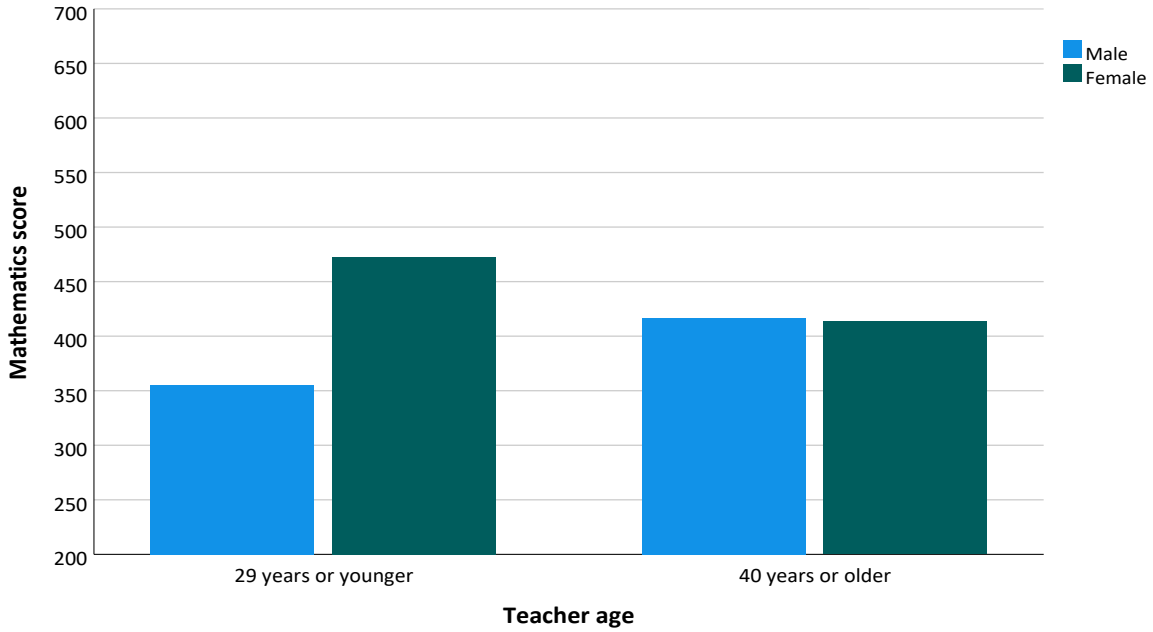
As shown in figures 4 and 5, a number of factors are associated with the underperformance of grade 4 boys in comparison to girls in school. One factor includes being absent from school once a week, which is more strongly associated with decreases in boys' mathematics achievement compared to girls'. This means that boys' mathematics achievement appears to suffer more from frequent absences from school compared to girls' achievement (figure 4). Teacher age also affects achievement of grade 4 boys and girls differently in Saudi Arabia. Having a younger teacher (29 years or younger), rather than an older teacher (40 years or older) is more strongly associated with lower mathematics achievement among boys compared to girls (figure 5).

Figure 4. Interaction between student gender and frequency of student absenteeism on mathematics achievement, grade 4



Note: The plot presents the predicted values from step 7 of the mathematics model for grade 4 (table 6). Only the reference category and the category for which a statistically significant interaction with student gender was found are presented.

Figure 5. Interaction between student gender and teacher age on mathematics achievement, grade 4



Note: The plot presents the predicted values from step 7 of the mathematics model for grade 4 (table 6). Only the reference category and the category for which a statistically significant interaction with student gender was found are presented.

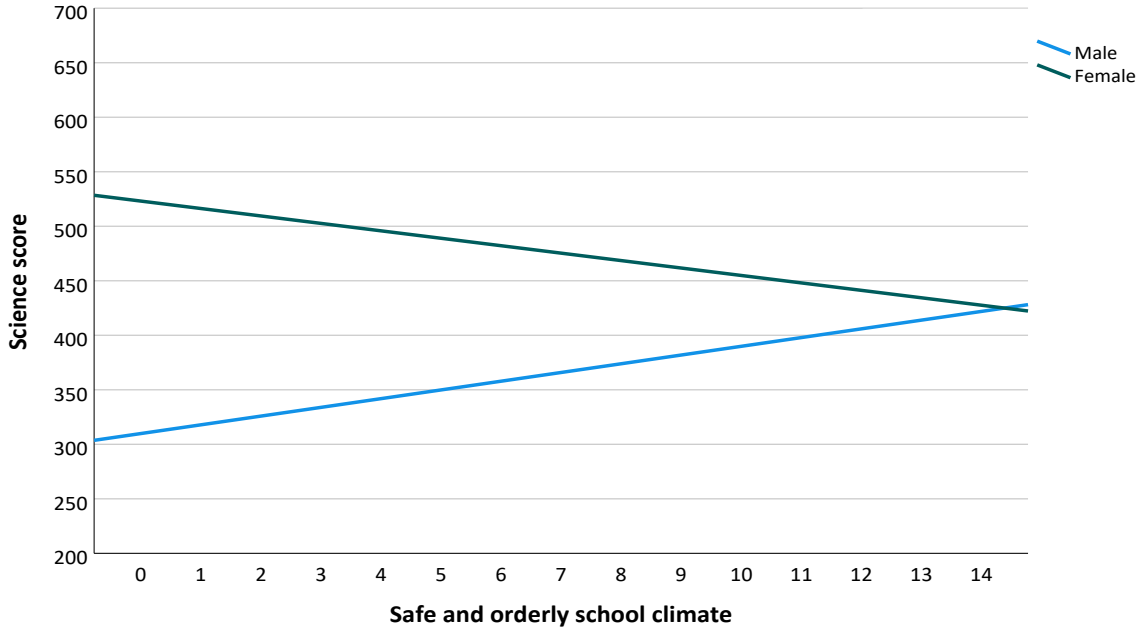
4.3.2 Science

Table 6 also presents the coefficients and model statistics for grade 4 science. Similar to grade 4 mathematics, the results from both steps 6 and 7 are presented. The final model explains 65 percent of the class/school-level variance and 27 percent of the student-level variance, or 39 percent of the total

observed variance. As shown in step 6, the coefficient of student gender is positive and statistically significant (B = 21.5, p < .05). This coefficient is much smaller in magnitude compared to the basic model (step 1 in table 5), which suggests that controlling for student, home, teacher, and school characteristics reduces the science achievement gap between grade 4 boys and girls by more than half. There is still, however, a significant unexplained gap between boys and girls in grade 4 science. Some potential explanations for this gap are discussed in section 5. Students' immigration status, early literacy and numeracy skills, absenteeism, bullying, attitudes toward science, and sense of school belonging are significantly associated with student achievement in science. For example, students who are absent once a week tend to underperform students who are never or almost never absent by 21 points. Also, students with stronger early literacy and numeracy skills tend to perform better in grade 4 science compared to other students (B = 7.6, p < .001).

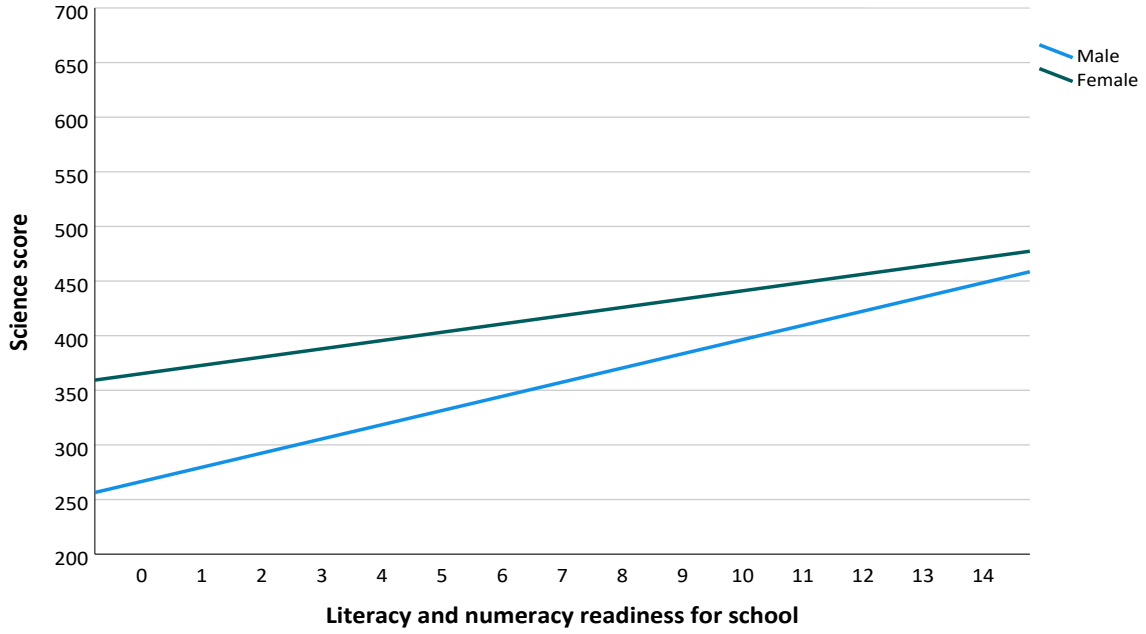
School location, school mean of home resources, teacher experience, age, and professional development are the class/school-level variables that were significantly associated with student performance in science. Surprisingly, after holding other variables constant, students in schools located in small towns or remote areas perform better in science than students in urban areas (B = 60.8, p < .01). Students of younger teachers scored lower in science than students of older teachers. Students whose teachers are between 30 and 39 years old scored 29 points lower than students of teachers who are 40 years or older and students whose teachers are 29 years old or younger scored 44 points lower than students of teachers who are 40 years or older. Also, students whose teachers have not participated in any professional development training on science (i.e., completed zero hours in professional development) tend to score much lower than students whose teachers have completed more than 35 hours in professional development in science (B = -35.3, p < .05). Consistent with grade 4 mathematics, results in science from step 7 show that poor school climate affects boys more negatively than girls (figure 6). The achievement gap between boys and girls is greater in schools with poor school climate relative to other schools. In addition, boys tend to benefit more from literacy and numeracy readiness than girls (figure 7), and girls in urban and suburban areas outperform boys (figure 8).

Figure 6. Interaction between student gender and safe and orderly school climate on science achievement, grade 4



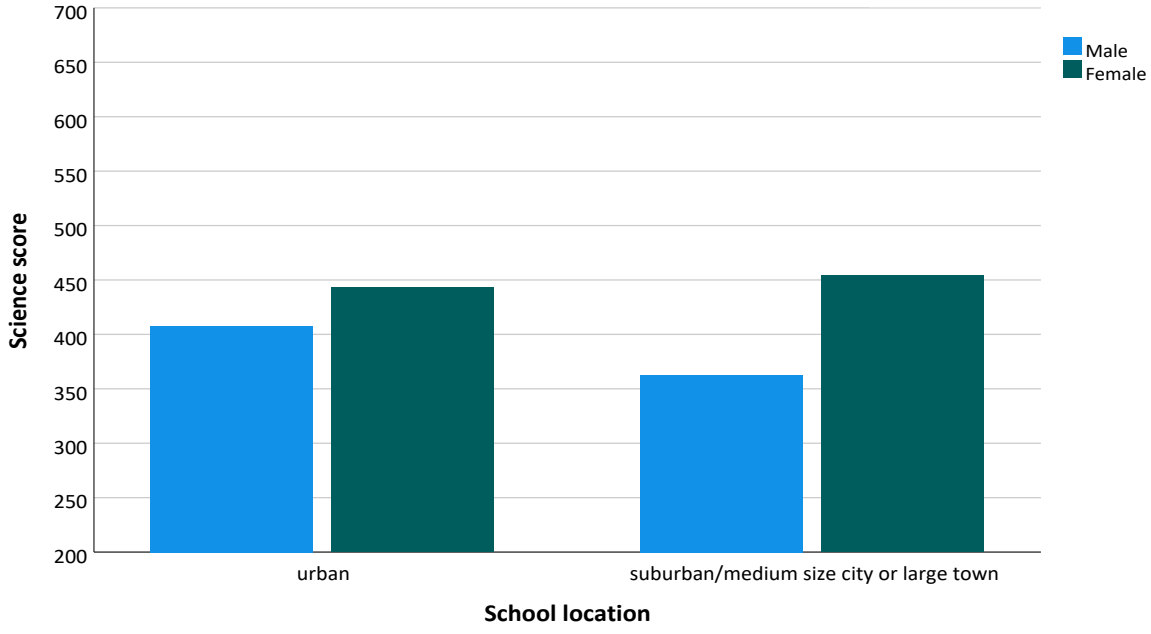
Note: The plot presents the predicted values from step 7 of the science model for grade 4 (table 6).

Figure 7. Interaction between student gender and literacy and numeracy readiness for school on science achievement, grade 4



Note: The plot presents the predicted values from step 7 of the science model for grade 4 (table 6).

Figure 8. Interaction between student gender and school location on science achievement, grade 4



Note: The plot presents the predicted values from step 7 of the science model for grade 4 (table 6). Only the reference category and the category for which a statistically significant interaction with student gender was found are presented.

Table 6. Steps 6 and 7 from the hierarchical two-level linear regression models for mathematics and science achievement, grade 4, TIMSS 2019

		Mathematics		Science	
		Step 6	**Step 7**	**Step 6**	**Step 7**
R^2	student-level (%)	21.0	26.2	22.5	27.4
	class/school-level (%)	54.9	71.4	46.0	64.8
Intercept (*SE*)		396.29 (21.00)	296.27 (37.86)	433.35 (21.20)	320.48 (35.66)
Student-level variables (reference category)		***B (SE)***	***B (SE)***	***B (SE)***	***B (SE)***
Student gender (male)		-6.71 (11.92)	164.98 (54.64)**	21.52 (8.79)*	249.50 (59.64)***
Student immigration status (native)					
	second-generation immigrant students	15.40 (11.36)	14.58 (11.44)	14.12 (11.98)	13.25 (12.08)
	first-generation immigrant students	26.74 (17.41)	26.53 (17.24)	32.10 (14.51)*	37.67 (13.70)**
Home resources for learning		4.35 (2.11)*	4.64 (2.11)*	2.85 (2.79)	2.83 (2.79)
Student owns mobile phone (no)		-0.10 (4.70)	0.10 (4.51)	-1.58 (1.96)	-1.45 (1.96)
Literacy and numeracy readiness for school		7.57 (1.44)***	7.57 (1.43)***	7.62 (1.41)***	10.35 (1.91)***
Preschool attendance and duration (did not attend)					
	1 year or less	1.51 (6.97)	0.38 (6.81)	4.88 (5.19)	3.79 (5.14)
	2 years	-6.20 (8.84)	-8.60 (8.84)	8.61 (6.40)	6.55 (6.36)
	3 years or more	15.10 (9.06)	14.00 (8.81)	8.71 (8.62)	8.66 (8.49)
Student absenteeism (never or almost never)					
	once every two months	5.36 (9.85)	4.74 (9.69)	7.90 (7.20)	8.14 (7.23)
	once a month	7.54 (8.60)	7.21 (8.49)	-2.91 (7.42)	-2.24 (7.39)
	once every two weeks	-20.34 (8.88)*	-21.87 (8.85)*	-15.47 (9.37)	-15.07 (9.20)
	once a week	-23.53 (7.35)**	-34.61 (8.67)***	-21.05 (5.67)***	-20.78 (5.50)***
Student likes learning mathematics/science		4.67 (1.60)**	4.82 (1.55)**	5.33 (1.76)**	5.16 (1.75)**
Student confident in mathematics/science		7.91 (1.50)***	8.24 (1.47)***	7.89 (1.62)***	8.09 (1.61)***
Student sense of school belonging		-3.10 (1.16)**	-3.10 (1.17)**	-3.68 (1.32)**	-3.55 (1.31)**
Bullying[a]		4.91 (1.28)***	5.20 (1.26)***	5.78 (1.44)***	5.91 (1.44)***
Class/school-level variables (reference category)					
School mean of home resources for learning		15.70 (9.03)	10.65 (8.07)	22.16 (6.74)**	17.82 (5.67)**
School location (urban)					
	suburban/medium size city or large town	-8.40 (11.43)	-3.81 (9.56)	-10.96 (9.66)	-36.03 (10.25)***
	small town or village/remote rural	54.91 (26.49)*	42.47 (24.77)	60.77 (20.36)**	60.10 (16.90)***
Teacher age (40 years or older)					
	29 years or younger	15.59 (19.55)	-4.73 (18.68)	-43.78 (17.99)*	-17.31 (16.16)
	30–39 years	16.37 (14.17)	18.35 (12.26)	-28.52 (10.26)**	-20.29 (9.24)*
Safe and orderly schools		0.32 (3.21)	6.63 (4.00)	5.29 (3.07)	9.92 (3.52)**
School emphasis on academic success		-0.26 (2.76)	2.93 (2.58)	-1.91 (1.83)	-0.75 (1.71)
School discipline		-3.91 (3.80)	-1.60 (3.74)	-1.84 (2.68)	-0.85 (2.51)
Teacher years of experience		1.76 (0.60)**	2.23 (0.59)***	-1.51 (0.60)*	-0.95 (0.59)

Class/school-level variables (reference category)	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*
Teacher job satisfaction	1.10 (4.34)	-0.03 (4.03)	-3.32 (3.34)	-4.52 (3.17)
Time assigned to mathematics/science homework (15 minutes or less)[b]	4.63 (8.85)	9.68 (8.52)	-7.30 (11.31)	-5.47 (9.81)
Poor teacher timekeeping (not a problem)				
minor problem	-24.39 (11.49)*	-14.89 (8.67)	2.50 (11.75)	4.86 (10.50)
moderate or serious problem	-2.18 (17.35)	13.44 (16.06)	-1.66 (17.19)	-0.39 (15.45)
Teacher absenteeism (not a problem)				
minor problem	8.04 (13.45)	5.03 (11.02)	1.01 (13.17)	-2.16 (10.94)
moderate or serious problem	-7.64 (23.90)	-8.53 (21.01)	1.87 (16.55)	5.55 (13.59)
Teacher major area of study (education and mathematics/science)				
mathematics/science but not education	-8.20 (9.26)	-6.76 (8.18)	-2.69 (9.78)	-7.35 (8.84)
all other majors	-9.88 (12.49)	-18.21 (10.74)	-23.80 (18.47)	-34.13 (17.72)
Professional development hours on mathematics/science (more than 35 hours)				
16–35 hours	-29.46 (11.69)*	-22.62 (10.50)*	3.50 (14.09)	5.25 (11.73)
6–15 hours	-5.35 (13.24)	4.99 (12.26)	-5.01 (12.01)	-4.52 (10.92)
less than 6 hours	-3.30 (14.44)	4.14 (12.28)	18.81 (15.70)	12.82 (13.31)
none	-1.71 (21.24)	-2.60 (16.21)	-35.32 (15.06)*	-37.77 (15.23)*
Professional development on mathematics/science content (no)	19.53 (10.85)	15.17 (10.60)	9.96 (12.76)	11.64 (11.14)
Professional development on mathematics/science pedagogy (no)	13.29 (11.26)	14.85 (9.46)	-2.53 (10.00)	-1.60 (9.00)
Principal years of experience	-0.13 (0.79)	-0.29 (0.71)	-0.20 (0.66)	0.15 (0.60)
Principal highest level of education (ISCED levels 7 & 8 — master's or doctorate degree)[c]	-3.76 (15.01)	4.78 (12.47)	-11.91 (10.93)	-13.19 (9.41)
Principal qualification in educational leadership (no)	-4.76 (13.02)	-6.06 (10.83)	5.72 (9.43)	7.28 (8.66)
School library (no)	19.09 (10.82)	15.10 (10.29)	4.69 (8.59)	7.65 (7.90)
Interaction terms (reference category)				
Student gender*Safe and orderly schools		-16.25 (5.13)**		-15.61 (4.93)**
Student gender*Student absenteeism (never or almost never) — once a week		22.88 (10.44)*		
Student gender*Teacher age (40 years or older) — 29 years or younger		89.46 (30.62)**		
Student gender*Literacy and numeracy readiness for school				-6.29 (2.57)*
Student gender*School location (urban) — suburban/medium size city or large town				75.76 (15.57)***
Loglikelihood (H_0)	-8549.92	-8533.11	-12381.76	-12361.03
Fit statistics *AIC*	17193.84	17166.23	24857.51	24822.05
BIC	17442.83	17431.12	25123.27	25104.77

Note: Null mathematics model: *Intercept* (*SE*): 402.02 (4.44), *ICC* = 0.31, H_0 = -32210.89, *AIC* = 64427.78, *BIC* = 64447.59. Null science model: *Intercept* (*SE*): 407.42 (4.88), *ICC* = 0.30, H_0 = -32704.43, *AIC* = 65414.86, *BIC* = 65434.67. *p < .05; **p < .01; ***p < .001. [a] Higher scores indicate less frequent bullying; [b] Other category: 16 minutes or more; [c] Other category: ISCED Level 6 — bachelor's or equivalent level.

4.4 Findings for grade 8

In this section, the results from the final two steps (steps 6 and 7) of each model are presented for grade 8 mathematics and science. Results for all steps are shown in appendix A3.

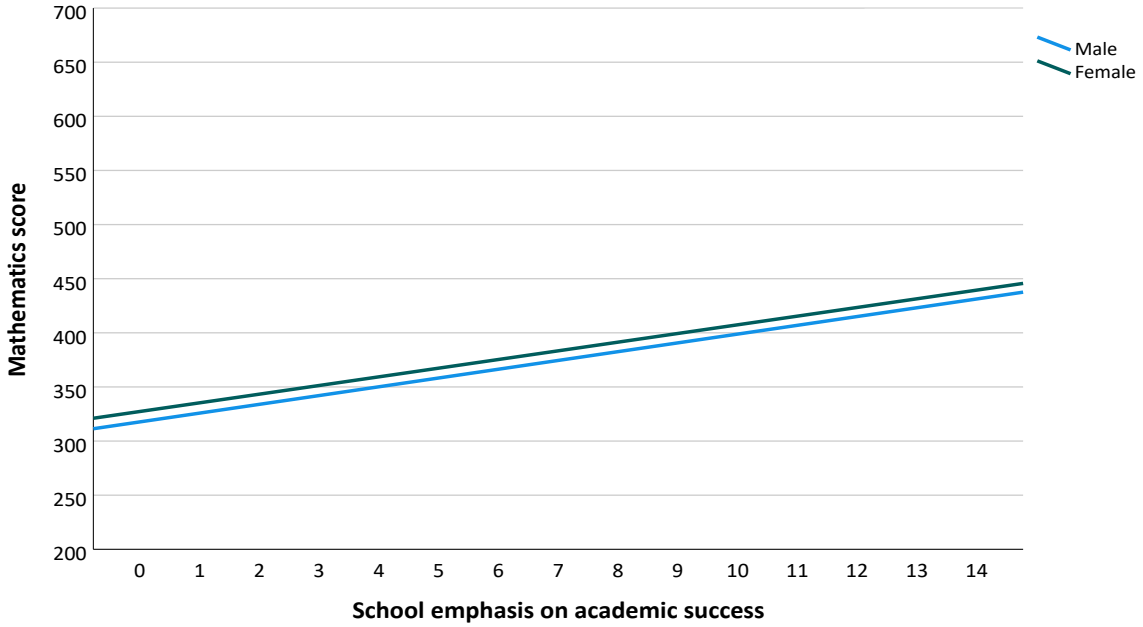
4.4.1 Mathematics

Table 7 presents the coefficients and model statistics for grade 8 mathematics. The gender difference in mathematics achievement remains statistically significant and only slightly smaller in magnitude (B = 16.3, p < .05) than the gender difference recorded in step 1 (B = 20.8, table 5), even after the addition of the selected conceptual blocks of predictor variables. Despite the relatively small effect of the predictor variables on the extent of the gender difference, the final model, including interactions, explains a substantial proportion of the observed variance in grade 8 mathematics achievement: 27 percent at level 1 and 74 percent at level 2, or 39 percent of the total observed variance.

Student-level factors that were significantly associated with higher mathematics achievement among both boys and girls were: first-generation or second-generation immigrant status, greater access to home learning resources, infrequent absence from school (no more than once every two months), greater confidence in mathematics, lower liking of mathematics, and taking less time to complete mathematics homework (15 minutes or less). School-level factors that were significantly associated with higher mathematics achievement among both boys and girls were: a higher school-average level of home resources for learning across the student body, and mathematics teachers having a postgraduate qualification (master's or doctorate) rather than a lower qualification.

One significant interaction with student gender was observed for grade 8 mathematics. This interaction, involving schools' emphasis on academic success, is illustrated in figure 9. The interaction term indicates that boys' mathematics achievement increases to a greater degree than girls' achievement in schools with stronger emphasis on academic success, relative to schools with a weaker emphasis on academic success. However, given that this difference was not substantial, this finding should be interpreted with caution.

Figure 9. Interaction between student gender and school emphasis on academic success on mathematics achievement, grade 8



Note: The plot presents the predicted values from step 7 of the mathematics model for grade 8 (table 7).

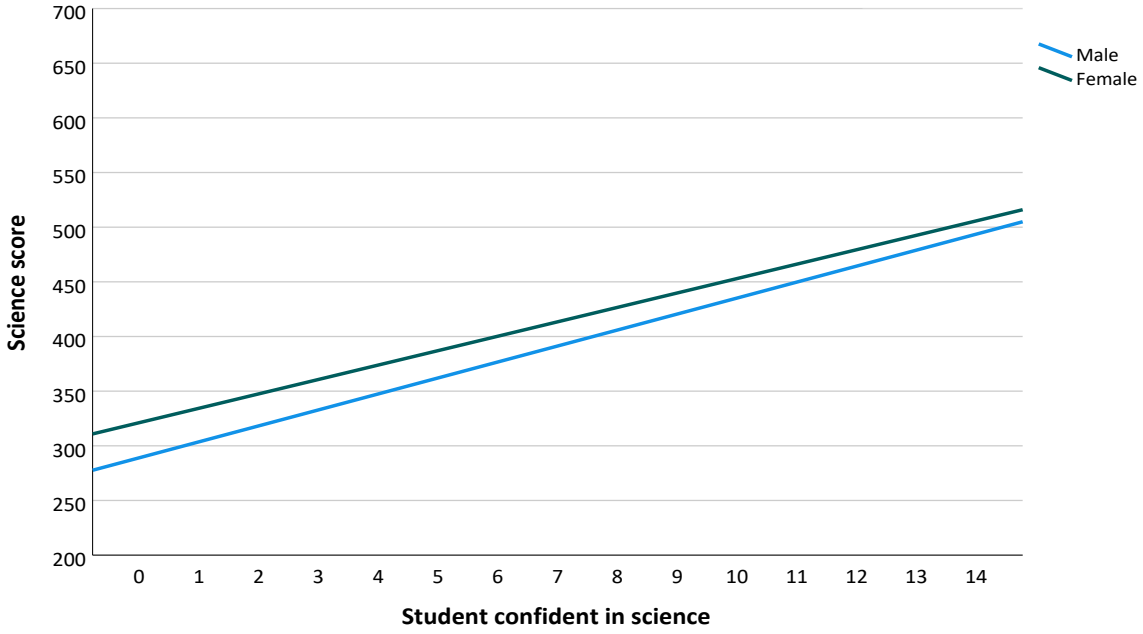
4.4.2 Science

Table 7 also presents the coefficients and model statistics for grade 8 science. The gender difference in science achievement remains statistically significant and substantial (B = 39.9, p < .001) after the addition of the selected conceptual blocks of predictor variables. Nonetheless, the final model, including interactions, explains a substantial proportion of the observed variance in grade 8 science achievement: 27 percent at level 1 and 79 percent at level 2, or 41 percent of the total observed variance.

Student-level factors that were significantly associated with higher science achievement among both boys and girls were: first-generation or second-generation immigrant status, greater access to home learning resources, infrequent absence from school (no more than once every two months), greater confidence in science, and taking less time to complete science homework (15 minutes or less). School-level factors that were significantly associated with higher science achievement among both boys and girls were: a higher school-average level of home resources for learning across the student body, infrequent teacher absenteeism (regarded by principals as not a problem), and science teachers whose qualification was in an area other than science or science education.

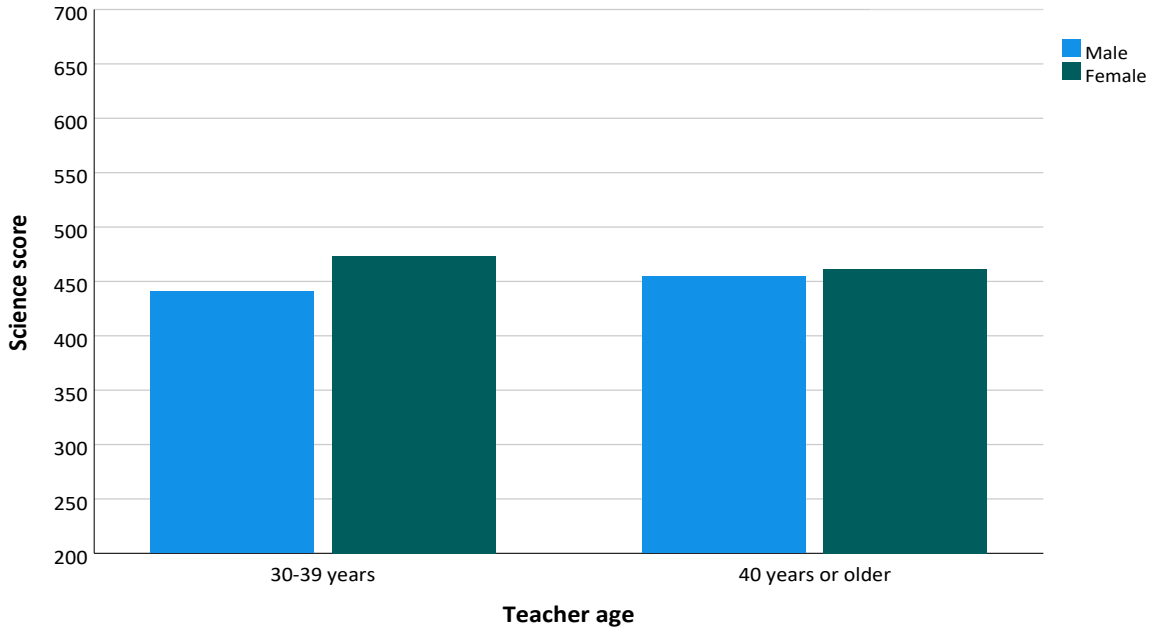
Two significant interactions with student gender were observed. These interactions, involving students' confidence in science and teachers' age, are illustrated in figures 10 and 11. The first interaction term indicates that higher levels of confidence in science are linked with greater gains in science achievement among grade 8 boys relative to grade 8 girls. The second interaction term indicates that boys' achievement in grade 8 science is higher when taught by older teachers, whereas girls' achievement is higher in classes taught by younger teachers (ages 30–39 years old) than in classes taught by older teachers.

Figure 10. Interaction between student gender and student confident in science on science achievement, grade 8



Note: The plot presents the predicted values from step 7 of the science model for grade 8 (table 7).

Figure 11. Interaction between student gender and teacher age on science achievement, grade 8



Note: The plot presents the predicted values from step 7 of the science model for grade 8 (table 7). Only the reference category and the category for which a statistically significant interaction with student gender was found are presented.

Table 7. Steps 6 and 7 from the hierarchical two-level linear regression models for mathematics and science achievement, grade 8, TIMSS 2019

		Mathematics		Science	
		Step 6	**Step 7**	**Step 6**	**Step 7**
R^2	student-level (%)	26.7	27.3	25.5	27.2
	class/school-level (%)	64.0	74.4	69.8	79.0
Intercept (*SE*)		367.09 (23.42)	310.38 (36.78)	441.38 (19.49)	426.88 (24.75)
Student-level variables (reference category)		***B (SE)***	***B (SE)***	***B (SE)***	***B (SE)***
Student gender (male)		16.29 (6.98)*	90.11 (35.49)*	39.89 (7.52)***	74.88 (25.07)**
Student immigration status (native)					
second-generation immigrant students		17.59 (7.92)*	17.51 (7.97)*	31.81 (8.36)***	31.84 (8.36)***
first-generation immigrant students		31.83 (9.37)**	31.92 (9.33)**	43.75 (9.67)***	43.76 (9.57)***
Home educational resources		4.09 (1.03)***	4.10 (1.03)***	5.39 (1.40)***	5.44 (1.40)***
Student owns mobile phone (no)		4.20 (5.13)	3.80 (5.07)	-0.32 (6.32)	-1.32 (6.53)
Student absenteeism (never or almost never)					
once every two months		-8.73 (5.81)	-9.10 (5.78)	-5.55 (6.26)	-5.16 (6.26)
once a month		-19.47 (5.06)***	-20.02 (5.20)***	-13.34 (5.34)*	-12.49 (5.24)*
once every two weeks		-29.92 (4.84)***	-30.31 (4.83)***	-10.40 (7.47)	-10.19 (7.49)
once a week		-44.23 (6.40)***	-44.52 (6.53)***	-46.77 (6.44)***	-47.10 (6.46)***
Student likes learning mathematics/science		-4.30 (1.57)**	-4.35 (1.57)**	-0.85 (1.17)	-0.79 (1.18)
Student confident in mathematics/science		15.24 (1.27)***	15.26 (1.27)***	11.35 (1.54)***	14.09 (2.02)***
Student sense of school belonging		-1.03 (1.12)	-1.00 (1.12)	-1.20 (1.45)	-1.15 (1.45)
Bullying[a]		-0.42 (0.91)	-0.43 (0.91)	1.49 (1.21)	1.32(1.21)
Time spent on mathematics/science homework (15 minutes or less)[b]		-11.35 (3.98)**	-11.40 (4.01)**	-12.37 (3.99)**	-11.94 (3.99)**
Class/school-level variables (reference category)					
School mean of home educational resources		25.97 (5.32)***	26.91 (5.29)***	20.23 (4.66)***	18.59 (4.37)***
School location (urban)					
suburban/medium size city or large town		6.64 (6.19)	9.63 (6.22)	-6.53 (5.96)	-7.84 (5.70)
small town or village/remote rural		2.05 (11.06)	4.58 (11.13)	3.12 (8.44)	-3.79 (8.53)
Teacher age (40 years or older)					
29 years or younger		-20.08 (19.13)	-23.36 (19.05)	9.78 (13.38)	3.56 (12.88)
30–39 years		-5.54 (11.15)	-6.51 (11.35)	-2.14 (6.47)	-20.34 (7.79)**
Safe and orderly schools		0.87 (1.68)	0.87 (1.62)	2.29 (1.40)	1.96 (1.44)
School emphasis on academic success		1.80 (1.70)	5.76 (2.48)*	0.07 (1.24)	0.46 (1.29)
School discipline		-2.60 (1.92)	-3.11 (1.93)	0.64 (1.14)	0.13 (1.13)
Teacher years of experience		-0.49 (0.88)	-0.60 (0.88)	0.54 (0.48)	0.54 (0.51)

Class/school-level variables (reference category)	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*
Teacher job satisfaction	-1.07 (2.18)	-1.70 (2.15)	-3.19 (2.47)	-3.35 (2.36)
Time assigned to mathematics/science homework (15 minutes or less)[b]	10.39 (6.03)	13.77 (6.21)*	-4.98 (6.32)	-7.03 (5.91)
Poor teacher timekeeping (not a problem)				
minor problem	7.38 (8.34)	7.72 (7.64)	2.15 (8.32)	0.82 (7.67)
moderate or serious problem	1.67 (15.21)	10.83 (14.07)	13.22 (11.46)	12.78 (10.64)
Teacher absenteeism (not a problem)				
minor problem	-3.46 (7.99)	-1.30 (7.43)	1.19 (6.79)	2.17 (6.24)
moderate or serious problem	-20.00 (13.91)	-28.83 (14.24)*	-17.19 (8.26)*	-16.62 (7.64)*
Teacher highest level of education (Up to ISCED level 6 — bachelor's or equivalent level)[c]	54.32 (15.71)**	67.81 (17.26)***	-3.91 (13.17)	-8.79 (12.29)
Teacher major area of study (mathematics/science and mathematics/science education)				
mathematics/science but not mathematics/science education	-5.55 (7.31)	-5.02 (7.25)	-11.49 (6.62)	-11.88 (6.43)
all other majors	-0.44 (9.16)	-5.21 (9.65)	22.96 (10.73)*	20.62 (9.38)*
Professional development hours on mathematics/science (more than 35 hours)				
16–35 hours	-14.62 (8.62)	-18.01 (8.67)*	-3.53 (7.63)	-3.66 (7.06)
6–15 hours	-0.91 (7.97)	-2.98 (7.67)	-0.71 (6.72)	-4.61 (6.57)
less than 6 hours	-9.99 (8.94)	-11.23 (8.69)	-0.04 (8.80)	3.27 (8.49)
none	-19.37 (11.70)	-24.04 (12.16)*	3.65 (15.01)	6.03 (14.52)
Professional development on mathematics/science content (no)	4.33 (5.31)	5.13 (5.06)	8.62 (7.78)	11.07 (7.34)
Professional development on mathematics/science pedagogy (no)	-0.02 (7.05)	1.28 (6.95)	-0.02 (8.96)	0.27 (7.98)
Principal years of experience	-0.24 (0.38)	-0.34 (0.39)	-0.09 (0.35)	-0.19 (0.35)
Principal highest level of education (ISCED levels 7 & 8 — master's or doctorate degree)[d]	-4.32 (11.75)	1.36 (11.86)	4.73 (13.27)	6.23 (13.36)
Principal qualification in educational leadership (no)	12.85 (8.06)	11.58 (7.68)	3.02 (6.33)	-0.05 (6.46)
School library (no)	9.32 (6.98)	9.76 (6.95)	6.66 (6.20)	9.22 (5.81)
Interaction terms (reference category)				
Student gender*School emphasis on academic success		-6.43 (3.04)*		
Student gender*Student confident in science				-4.88 (1.91)*
Student gender*Teacher age (40 years or older) — 30–39 years				32.15 (10.20)**
Loglikelihood (H_0)	-12113.38	-12109.54	-13016.61	-13004.99
Fit statistics *AIC*	24316.77	24311.07	26123.23	26103.99
BIC	24572.77	24572.76	26381.44	26373.68

Note: Null mathematics model: *Intercept* (*SE*): 394.83 (2.88), *ICC* = 0.25, H_0 = -32297.78, *AIC* = 64601.56, *BIC* = 64621.49. Null science model: *Intercept* (*SE*): 431.76 (3.69), *ICC* = 0.27, H_0 = -33003.98, *AIC* = 66013.96, *BIC* = 66033.89. *p < .05; **p < .01; ***p < .001. [a] Higher scores indicate less frequent bullying; [b] Other category: 16 minutes or more; [c] Other category: ISCED levels 7 & 8 —master's or doctorate degree; [d] Other category: ISCED Level 6 — bachelor's or equivalent level.

5. Discussion

The results presented in the previous section shed some light on the factors that are associated with mathematics and science achievement in Saudi Arabia, and on the factors that contribute to the large observed differences in achievement between boys and girls. Although there was variation across the four sets of multilevel models in terms of which variables were associated with student achievement when considered simultaneously, some consistency was also evident. Such consistency should help identify key factors for educators and policy makers in Saudi Arabia to consider as part of broader efforts to raise levels of achievement in elementary and intermediate schools.

Summarized below are the most important findings of the analyses presented in the results section and some of the broader issues arising from these findings. In particular, attention is drawn to the most robust findings with the clearest implications for educators and policy makers in Saudi Arabia.

5.1 Summary of main findings

The results of the analysis described in this paper suggest that, at the elementary level, early literacy and numeracy skills, student absenteeism, and school climate contribute to the observed gender gap in student performance in Saudi Arabia. A significant interaction was noted between student gender and students' early literacy and numeracy readiness for school in predicting grade 4 science achievement. Specifically, boys' science achievement in grade 4 was particularly low, relative to girls', among students with poorly developed preschool literacy and numeracy skills (as reported by their parents). In addition, the findings showed that boys in grade 4 appear to be particularly disadvantaged by disorderly or unsafe school environments. Substantial gender differences in favor of girls among students attending disorderly schools, in both mathematics and science, are reduced to minor differences among students attending highly safe and orderly schools. Similarly, while all four sets of models showed a negative association between frequent student absenteeism and achievement, the association was particularly strong among grade 4 boys.

Overall, several of the variables examined were found to be significantly associated with both mathematics and science achievement in grade 4, for both boys and girls. Higher scores in mathematics and science were associated with several factors, including students: (a) having stronger early literacy and numeracy skills upon starting primary school, (b) liking mathematics or science, (c) being more confident in mathematics or science, (d) being present at school more regularly, and (e) having a lower frequency of bullying. The set of factors most consistently associated with achievement at grade 4 are predominantly at the student level and drawn mostly from the first two conceptual blocks entered into the models: the home background and student engagement and attitudes.

Similarly, several variables were found to be significantly associated with both mathematics and science achievement in grade 8. However, the pattern of common variables is somewhat different between the two grade levels. Among grade 8 students, higher scores in mathematics and science were associated with students': (a) immigration status, (b) access to more learning resources at home, (c) higher confidence in mathematics or science, (d) regular presence in school, and (e) enrollment in a school where students have a higher average level of home learning resources. Similar to the findings at grade 4, each of these variables was part of the first two conceptual blocks in the models (the home background and student engagement and attitudes), with 4 of the 5 being student-level factors.

5.2 Accounting for observed gender differences in achievement in Saudi Arabia

Student gender remained a significant predictor of science achievement in grade 4, and both mathematics and science achievement in grade 8, even after accounting for the effects of the other predictors. Although the gender difference in achievement is partially accounted for by the modeled variables—leading to a reduction in the 'remaining' or residual gender difference in all models—grade 4

mathematics was the only one of the four sets of models where the final gender difference was no longer statistically significant. This implies that other factors, not examined in the models, contribute to the substantial residual gender difference in grade 4 science and grade 8 mathematics and science. One possibility is that selection effects could be driving the observed differences — that is, if only high-achieving girls attend school or sit for assessments, but most boys do so, there is a possibility of bias, such that girls' average achievement would appear inflated. However, as enrollment in primary and intermediate education in Saudi Arabia is almost universal among both genders, selection effects are unlikely to be playing a role in this analysis.

Given evidence from other studies, it is likely that differences in reading proficiency play a role in explaining at least part of the remaining gender differences. This is particularly so in relation to science achievement, where test items are by necessity embedded in a context that often requires a greater degree of reading comprehension. Differences in reading proficiency between boys and girls might also contribute to explaining the remaining gender differences observed here in mathematics achievement in grade 8, given that grade 8 mathematics assessments require more reading skills compared to mathematics in early grades. For example, test items assessing applied reasoning or problem-solving skills in grade 8 are more likely to be embedded in a short scenario requiring some level of reading. International assessments in recent years have consistently shown that reading achievement is, on average, substantially lower in Saudi Arabia than in many other countries, both among grade 4 students (Mullis et al. 2012; 2017) and among 15-year-old students (OECD 2019).

Moreover, in Saudi Arabia, the reading achievement gap in which girls outperform boys is among the largest gender differences in the world. Differences between boys and girls in reading proficiency have been found to exceed half a standard deviation in both the PIRLS and PISA studies (Mullis et al. 2017; OECD 2019). Previous research on TIMSS mathematics and science items has shown that items with a higher reading load (those requiring a higher volume of reading or more complex reading skills) tend to be more difficult for students to answer correctly than items with a lower reading load, and also that weaker readers tend to be disproportionately disadvantaged by a higher reading load (Mullis, Martin, and Foy 2013). For this reason, the magnitude and consistency of Saudi boys' relative disadvantage in reading, seen across various studies, seems likely to play a role in contributing to their poorer results found here in mathematics and science achievement even after accounting for a range of contextual variables. The NALO 2018 results provide further support for this view, with boys at both grade levels being more likely than girls to report not having read a book (although it should be noted that this was the case even for a substantial minority of girls).

The proposed importance of reading skills in underpinning mathematics and scientific achievement is consistent with the pattern of residual (unexplained) variance reported in the models, which indicates a role for other factors operating largely at the student level. After accounting for a range of other student- and class/school-level factors, the majority (approximately three-quarters) of class/school-level variance was explained in the models, whereas a majority of student-level variance remained unexplained. This suggests that the residual gender differences in achievement are likely to be associated more strongly with student-level factors, such as reading skills, social and behavioral skills, or aspects of the home background, than with additional school-level factors. The finding from NALO 2018 that more boys than girls have repeated a grade at school because of poor academic performance is worth noting in this regard. Similar data on the extent of grade repetition are available from PISA 2018 (OECD 2019), where 13.0 percent of 15-year-old boys in Saudi Arabia reported repeating at least one grade, compared to 9.8 percent of girls. These findings hint at the likelihood that early disadvantages and difficulties with learning in the early grades may compound over time, and that there is a need for stronger learning supports for students with special educational needs to enable progression through the education system. Although this issue affects both boys and girls in Saudi Arabia, the figures from NALO and PISA indicate that such compounding educational disadvantage is more clearly apparent among boys.

Teaching quality is another factor that appears to be associated with gender differences in achievement in Saudi Arabia. Female entrants to the teaching profession in Saudi Arabia tend to score higher than their male counterparts on the teacher licensure examination. This is consistent with evidence from other countries, which shows that the teaching profession attracts more high-ability female teachers than male teachers (Carroll, Parasnis, and Tani 2021; Corcoran, Evans, and Schwab 2004). The differences in abilities between female and male teachers could be explained, in part, by the gender differences in returns to education across occupations (World Bank 2012; Cortes and Pan 2018). Research on teacher labor markets has shown that the opportunity cost of becoming a teacher is lower for women than men, due primarily to the limited occupational opportunities for women outside the field of education (Carroll, Parasnis, and Tani 2021). Additionally, teaching is traditionally a preferable profession for women in Saudi Arabia. The competition among female graduates for teaching jobs in Saudi Arabia is much higher than the competition among males. From the demand side, this implies a higher probability of selecting cognitively talented teachers from female graduates than from male graduates.

The analyses presented in this paper have accounted for a high proportion of the observed variance in mathematics and science achievement in Saudi Arabia (ranging from 39 percent to 41 percent across subject domains and grade levels). Notably, these models largely account for the portion of variance in achievement that can be attributed to the class/school level. This suggests that policy makers may reasonably hope that focusing their attention on improving the class/school-level issues identified here (e.g., safe and orderly school climate, support for academic achievement, teacher attendance and timekeeping) would contribute to creating an education system that promotes higher levels of student achievement. However, the fact that the majority of variance in achievement (70–75 percent) is attributable to student-level factors means that policy makers will also have to look at the home environment and broader society, as well as the school environment, in order to raise levels of achievement and close the (currently very wide) gaps in achievement between boys and girls in Saudi Arabia. This student-level variance has been partly explained by variables examined here (e.g., students' early literacy and numeracy skills, attitudes toward mathematics and science, experiences of bullying, and level of access to home resources for learning), but significant variance remains unexplained.

5.3 Limitations

The conclusions that can be drawn from these analyses are limited to being correlational in nature, as TIMSS and NALO are both cross-sectional studies. This means that it would be incorrect to claim on the basis of these results alone that changes in any of the included variables will lead to corresponding changes in student achievement. In some cases, the findings presented here are clearly consistent with theoretical expectations and evidence from other settings—for example, that promoting more regular attendance at a school with a learning-supportive climate may reasonably be expected to have positive implications for student learning. Nonetheless, readers should be aware that the model results need to be interpreted cautiously and with due regard to the wider theoretical and empirical literature. Informed decisions should be based on a broad reading of the literature and the evidence base, including the new results presented in this paper, rather than on any single study.

The results of the models hint at the importance of teachers and teaching quality as contributing factors to student outcomes. However, the strength of any conclusions drawn in this paper related to teaching are constrained by limitations in the available data. For example, the TIMSS variable describing teachers' qualifications at grade 4 was omitted from analysis due to an error identified in the Saudi Arabia data set for TIMSS 2019. TIMSS collects some data related to teachers and classroom practices but more detailed analysis on teacher quality would be possible with research studies and, thus, data sets more focused on this topic.

Finally, although the highly gender-segregated structure of the education system in Saudi Arabia presents an opportunity to examine the educational environments experienced by boys and girls in

relative isolation, this same feature also imposes analytic constraints. As there are no cases in the available data of boys and girls taught in the same classes, boys taught by female teachers, or girls taught by male teachers, it is impossible to disentangle gendered differences in learning outcomes from other factors that covary completely with students' gender. For example, in these data sets boys in Saudi Arabia are universally taught by male teachers, who, in turn, received their education and teaching qualifications from all-male institutions, which may differ in important ways from the institutions attended by girls and female teachers. The ongoing rollout of a scheme to assign female teachers to boys in the early grades, as described earlier, will provide opportunities in future to reexamine outcomes among boys and girls while controlling for teacher characteristics to a greater degree.

6. Conclusions and Implications

The findings of this study point to the relevance of the school climate in understanding the current gender differences in achievement observed in Saudi Arabia. Although previous research points to the value to students of a stable and supportive school climate in general (Nilsen et al. 2016; Reynolds et al. 2014), the results of this study indicate that boys in Saudi Arabia may be especially impacted by a negative or unstable school environment. Most notably, the presence of a safe and orderly school climate for grade 4 students, and a supportive climate for academic success for grade 8 students, are particularly associated with higher achievement for boys relative to attending less orderly or less supportive schools.

School principals, teachers, and other educators should be cognizant of the importance of these factors and should take active steps to build and maintain positive school and classroom environments where students feel safe, connected, and positively challenged to learn and think. Where these conditions are not present, student learning is likely to be impeded. This is especially the case for boys, who may require a greater degree of behavioral support and guidance from adults to engage fully with schoolwork in a structured classroom setting in a single-gender school environment. Where such support and guidance are lacking, boys appear to fall behind in their learning and are at risk of being held back for a year to a greater degree than girls who similarly lack a positive school climate. This may be related to gendered differences in societal expectations (Ridge and Jeon 2020) and, as indicated by the NALO data, greater support for learning for girls at home (Ridge and Jeon 2020). Teachers can help to create positive learning environments and encourage active student participation in their learning by, for example, integrating students' interests into the lesson material where possible and remaining alert to the effects of stereotypes, such as boys being more suited than girls to science and mathematics, and girls being more suited to reading, that may negatively affect how students engage with lessons and how teachers communicate with their students (Brozo et al. 2014; OECD 2015). It is also important that lessons are challenging but at a level that students can realistically engage with and understand; where basic prerequisite learning has not been solidified, teachers are likely to find themselves covering more advanced topics with limited student engagement or understanding (Niemiec and Ryan 2009). Other practices that teachers can integrate into their teaching in order to create a positive learning environment include offering students choices, providing rationales for decisions made or where choices cannot be offered, encouraging students to ask questions and to offer their perspectives, listening to and acknowledging students' contributions, and offering constructive feedback on how students can improve (Teixeira et al. 2020).

A supportive school environment is important for boys' learning, but support for learning in the home is also crucial. By the time students begin attending school, they have been growing, developing, and learning at home and in the community for several years already. The TIMSS data show that boys in Saudi Arabia tend to begin school with weaker early literacy and numeracy skills than girls. More than that, the models indicate that boys' science achievement is more strongly associated with their early literacy and numeracy skills compared to girls'. Boys who begin school with weak early literacy and

numeracy skills tend to have considerably lower science achievement than their female counterparts with equivalent early literacy and numeracy skills by grade 4, while science achievement of boys and girls with stronger early literacy and numeracy skills tends to be similar. In other words, boys who begin school at an early learning disadvantage to their peers are further disadvantaged as they progress through the education system and appear to be at more risk of falling behind than girls who begin school with weaker early skills. This can also be seen in students' reports, in NALO, that boys are more likely to repeat a year in school because of poor academic performance.

It is important that parents are aware that early childhood development lays a foundation for future education, health, well-being, and economic success. Public health and education agencies should promote awareness among parents and provide guidance and resources to encourage greater engagement in early learning in the family. For example, simple activities that can contribute to a child's early literacy and numeracy development could include reading together, describing a scene in everyday life, counting everyday objects or singing counting songs, and using mathematical and spatial language while playing with shapes or other objects (e.g., "behind", "above", "beside", "straight", "curved", "double"). Data from TIMSS 2019 show that parents in Saudi Arabia report engaging in activities of these types less frequently with young boys than young girls. Taking steps to increase the level of support for early childhood learning at home for boys, in particular, would likely lead to a stronger foundation in the future for boys starting school and to greater progress in learning among boys. Cultural and social barriers that contribute to low enrollment of young children in kindergarten—for example, social expectations relating to motherhood and childrearing at home—also need to be considered in this respect.

It is noteworthy that, despite the differences seen across the four sets of models, two variables were found to be significantly associated with both mathematics and science achievement at both grade levels. These were students' confidence in mathematics or science (positively associated with achievement in all cases) and students' reported levels of absenteeism (more frequent absences being negatively associated with achievement in all cases). The consistency of these findings demands attention from Saudi Arabia's education community. In particular, student absenteeism, as an issue that is likely more responsive to policy making compared to others (e.g., students' socio-economic status), should be considered carefully. The analyses presented in this paper have shown that student absenteeism in Saudi Arabia is widespread, frequent, and consistently associated with achievement in at least two key areas of study (mathematics and science), at both elementary and intermediate school levels. In many countries, student absenteeism is relatively rare and structures are in place to monitor and promote regular attendance at school. These structures can encompass both informal channels (e.g., between the schoolteacher or principal and the child's parents) and formal channels (e.g., formal communication between the school and the home or, in more extreme cases, a state agency tasked to follow up to ensure minimal levels of attendance at school). The frequency of absenteeism for many students in Saudi Arabia, coupled with the likely negative implications of regular absenteeism for achievement, suggest that Saudi Arabia's policy makers should study efforts in other countries to combat absenteeism (e.g., Knoster 2016; Rogers and Vegas 2009) and consider how similar approaches could be usefully adapted to the local context.

A similar problem is apparent with the teaching workforce in Saudi Arabia's schools. A substantial proportion of school principals, at both elementary and intermediate levels, indicated that teacher absenteeism and poor teacher timekeeping (teachers arriving late to school or leaving early) are problems in their schools. This is consistent with previous research indicating that teachers in Saudi Arabia's schools often lack enthusiasm for the profession and are poorly motivated (OECD 2020).. Without taking steps to ensure that teachers are both highly skilled and present and engaged in teaching during scheduled working hours, students will continue to be at risk of failing to reach their full potential as a result of failures in school management practices. Other initiatives that may be taken to, for example, build supportive school climates, are likely to be limited as long as they are undermined by poor teacher

attendance at school and lack of teacher enthusiasm (in itself, a contributory factor to a school environment that is not conducive to student learning).

Teacher training represents another area for improvement. Results from this paper show that despite male teachers' greater exposure to education during initial training and their higher qualifications compared to female teachers,[10] boys in Saudi schools achieve much poorer outcomes than girls in both mathematics and science. Although holding higher qualifications does not necessarily suggest a higher standard of teaching or improved student performance (Harris and Sass 2011), especially when the focus of the qualification is unknown, these patterns may signal the poor quality of teacher education and training. Further study of these dynamics as they relate to student outcomes would be useful.

In general, efforts to raise educational achievement in Saudi Arabia require taking a broader view beyond the necessary focus on schools and teachers. As noted above, early child development (physical, cognitive, social) and early learning provide foundations for achievement in elementary and intermediate school, and beyond. Ongoing support for learning at home throughout childhood is also crucial, including modeling of positive behaviors (e.g., reading) and involvement in children's education by their parents.

6.1 Suggestions for further research

Future efforts to explain the observed differences between boys' and girls' achievement in Saudi Arabia should, if possible, seek to include a broader range of out-of-school factors in the analysis than were possible with the TIMSS 2019 data set. For example, some variables that are not available in TIMSS 2019 or NALO 2018 but that could be usefully considered in a future analysis of gender differences in achievement include students' reading proficiency; students' engagement in reading for leisure; and the (gendered) nature of parents' expectations and aspirations for their child's education, qualifications, and future career. In particular, considering the importance of literacy as a foundational skill (see also Gregory et al. 2021), the inclusion of an indicator of reading achievement would help control for gender differences relating to literacy levels and would allow more fine-grained examination of mathematical and scientific proficiency. Among international assessments, data from PISA at intermediate level or from a joint TIMSS and PIRLS assessment (such as TIMSS/PIRLS 2011) at primary level could be used for this purpose. At the national level, an administration of NALO that assessed reading as well as mathematics or science from the same students could also be used. Given that the majority of unexplained variance in the models presented in this paper was at the student level, extending future analyses in this way should provide further useful insights.

As noted above (section 5.3), a focused examination of teaching quality in Saudi Arabia — incorporating teacher characteristics, quality of teacher education, professional development, availability and use of resources, classroom management, professional collaboration, and pedagogy — would shed further light on some of the points raised in this paper. In particular, differences between the classroom environments of boys and girls, given the gender-segregated structure of the education system, merit closer inspection.

Finally, it would be useful to extend the work presented in this paper by drawing on data from other countries. In the first instance, subsequent research could focus on countries with similar cultural contexts such as other countries with comparable international data within the Gulf or MENA regions. Such research could examine (a) the extent of similarity between observed gender differences in Saudi Arabia compared to other countries, and (b) similarities and differences in the factors associated with student outcomes in each national context. Further work could also usefully examine factors associated with gender differences in single-gender compared to mixed-gender schools (see figure 2), particularly in, but not limited to, the MENA region.

[10] Male teachers are more likely than female teachers to report holding a master's or doctorate-level qualification, as are principals of boys' schools compared to principals of girls' schools.

Acknowledgments

The authors would like to thank Harry Patrinos, Georgios Sideridis, and Tarek Mostafa for their valuable comments on a draft version of this paper. The team is also grateful to Laura Gregory, Andreas Blom, and Yisgedullish Amde for their helpful feedback and support. Jee Yoon Lee edited the paper.

References

Angrist, Noam, Simeon Djankov, Pinelopi K. Goldberg, and Harry A. Patrinos. 2021. "Measuring Human Capital Using Global Learning Data." *Nature* 592: 403–8. https://doi.org/10.1038/s41586-021-03323-7.

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman. 2016. "School Quality and the Gender Gap in Educational Achievement." *American Economic Review* 106 (5): 289–95. https://doi.org/10.1257/aer.p20161074.

———. 2019. "Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes." *American Economic Journal: Applied Economics* 11 (3): 338–81. https://doi.org/10.1257/app.20170571.

Bertrand, Marianne, and Jessica Pan. 2013. "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior." *American Economic Journal: Applied Economics* 5 (1): 32–64. https://doi.org/10.1257/app.5.1.32.

Brozo, William G., Sari Sulkunen, Gerry Shiel, Christine Garbe, Ambigapthy Pandian, and Renate Valtin. 2014. "Reading, Gender, and Engagement." *Journal of Adolescent & Adult Literacy* 57 (7): 584–93. https://doi.org/10.1002/jaal.291.

Buchmann, Claudia, and Thomas A. DiPrete. 2006. "The Growing Female Advantage in College Completion: The Role of Family Background and Academic Achievement." *American Sociological Review* 71 (4): 515–41. https://doi.org/10.1177/000312240607100401.

Carroll, David, Jaai Parasnis, and Massimiliano Tani. 2021. "Why Do Women Become Teachers While Men Don't?" *The B.E. Journal of Economic Analysis & Policy* 21 (2): 793–823. https://doi.org/10.1515/bejeap-2020-0236.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Corcoran, Sean P., William N. Evans, and Robert M. Schwab. 2004. "Women, the Labor Market, and the Declining Relative Quality of Teachers." *Journal of Policy Analysis and Management* 23 (3): 449–70. https://www.jstor.org/stable/3326261.

Cornwell, Christopher, David B. Mustard, and Jessica Van Parys. 2013. "Noncognitive Skills and the Gender Disparities in Test Scores and Teacher Assessments: Evidence from Primary School." *Journal of Human Resources* 48 (1): 236–64. https://doi.org/10.3368/jhr.48.1.236.

Cortes, Patricia, and Jessica Pan. 2018. "Occupation and Gender." In *The Oxford Handbook of Women and the Economy*, edited by Susan L. Averett, Laura M. Argys, and Saul D. Hoffman, 424–52. New York: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190628963.013.12.

Davier, Matthias von, Eugenio Gonzalez, and Robert J. Mislevy. 2009. "What Are Plausible Values and Why Are They Useful?" *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments* 2: 9–36. https://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_02_Chapter_01.pdf.

DiPrete, Thomas A., and Claudia Buchmann. 2013. *The Rise of Women: The Growing Gender Gap in*

Education and What It Means for American Schools. New York: Russell Sage Foundation.

DiPrete, Thomas A., and Jennifer L. Jennings. 2012. "Social and Behavioral Skills and the Gender Gap in Early Educational Achievement." *Social Science Research* 41 (1): 1–15. https://doi.org/10.1016/j.ssresearch.2011.09.001.

Downey, Douglas B., and Anastasia S. Vogt Yuan. 2005. "Sex Differences in School Performance during High School: Puzzling Patterns and Possible Explanations." *The Sociological Quarterly* 46 (2): 299–321. https://doi.org/10.1111/j.1533-8525.2005.00014.x.

Fortin, Nicole M., Philip Oreopoulos, and Shelley Phipps. 2015. "Leaving Boys Behind: Gender Disparities in High Academic Achievement." *Journal of Human Resources* 50 (3): 549–79. https://doi.org/10.3368/jhr.50.3.549.

Fritz, Catherine O., Peter E. Morris, and Jennifer J. Richler. 2012. "Effect Size Estimates: Current Use, Calculations, and Interpretation." *Journal of Experimental Psychology: General* 141 (1): 2–18. https://doi.org/10.1037/a0024338.

Gregory, Laura, Hanada Taha Thomure, Amira Kazem, Anna Boni, Mahmoud A. A. Elsayed, and Nadia Taibah. 2021. *Advancing Arabic Language Teaching and Learning — A Path to Reducing Learning Poverty in the Middle East and North Africa*. Washington, DC: World Bank. https://documents1.worldbank.org/curated/en/909741624654308046/pdf/Advancing-Arabic-Language-Teaching-and-Learning-A-Path-to-Reducing-Learning-Poverty-in-the-Middle-East-and-North-Africa.pdf.

Harris, Douglas N., and Tim R. Sass. 2011. "Teacher Training, Teacher Quality and Student Achievement." *Journal of Public Economics* 95 (7–8): 798–812. https://doi.org/10.1016/j.jpubeco.2010.11.009.

Hattie, John A. C. 2009. *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement*. London, New York: Routledge.

Jha, Jyotsna, and Sarah Pouezevara. 2016. *Measurement and Research Support to Education — Strategy Goal I — Boys' Underachievement in Education: A Review of the Literature with a Focus on Reading in the Early Years*. Washington, DC: United States Agency for International Development (USAID).

Knoster, Kevin C. 2016. *Strategies for Addressing Student and Teacher Absenteeism: A Literature Review*. Washington, DC: U.S. Department of Education, North Central Comprehensive Center.

LaRoche, Sylvie, Marc Joncas, and Pierre Foy. 2020. "Sample Design in TIMSS 2019." In *Methods and Procedures: TIMSS 2019 Technical Report*, edited by Michael O. Martin, Matthias von Davier, and Ina V. S. Mullis, 3.1-3.33. TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College, and International Association for the Evaluation of Educational Achievement (IEA). https://timssandpirls.bc.edu/timss2019/methods/chapter-3.html.

Legewie, Joscha, and Thomas A. DiPrete. 2012. "School Context and the Gender Gap in Educational Achievement." *American Sociological Review* 77 (3): 463–85. https://doi.org/10.1177/0003122412440802.

Mullis, Ina V. S., Michael O. Martin, and Pierre Foy. 2013. "The Impact of Reading Ability on TIMSS Mathematics and Science Achievement at the Fourth Grade: An Analysis by Item Reading Demands." In *TIMSS and PIRLS 2011: Relationships among Reading, Mathematics, and Science Achievement at the Fourth Grade —Implications for Early Learning*, edited by Michael O. Martin and Ina V. S. Mullis, 67–108. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, and International Association for the Evaluation of Educational Achievement (IEA). https://timssandpirls.bc.edu/timsspirls2011/downloads/TP11_Chapter_2.pdf.

Mullis, Ina V. S., Michael O. Martin, Pierre Foy, and Kathleen T. Drucker. 2012. *PIRLS 2011 International Results in Reading*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston

College, and International Association for the Evaluation of Educational Achievement (IEA). https://doi.org/10.1097/01.tp.0000399132.51747.71.

Mullis, Ina V. S., Michael O. Martin, Pierre Foy, and Martin Hooper. 2017. *PIRLS 2016 International Results in Reading*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, and International Association for the Evaluation of Educational Achievement (IEA).

Mullis, Ina V. S., Michael O. Martin, Pierre Foy, Dana L. Kelly, and Bethany Fishbein. 2020. *TIMSS 2019 International Results in Mathematics and Science*. TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College, and International Association for the Evaluation of Educational Achievement (IEA).Muthén, Linda K., and Bengt O. Muthén. 2017. "Mplus User's Guide." 8th ed. Los Angeles, CA: Muthén and Muthén.

Niemiec, Christopher P., and Richard M. Ryan. 2009. "Autonomy, Competence, and Relatedness in the Classroom: Applying Self-Determination Theory to Educational Practice." *Theory and Research in Education* 7 (2): 133–44. https://doi.org/10.1177/1477878509104318.

Nilsen, Trude, Sigrid Blömeke, Kajsa Yang Hansen, and J. E. Gustafsson. 2016. "Are School Characteristics Related to Equity? The Answer May Depend on a Country's Developmental Level." Policy Brief 10, International Association for the Evaluation of Educational Achievement (IEA), Amsterdam, Netherlands.

OECD (Organisation for Economic Co-operation and Development). 2015. *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*. PISA. Paris: PISA, OECD Publishing. https://doi.org/10.1787/9789264229945-en.

———. 2019. *PISA 2018 Results (Volume I): What Students Know and Can Do*. Paris: PISA, OECD Publishing. https://doi.org/10.1787/5f07c754-en.

———. 2020. *Reviews of National Policies for Education: Education in Saudi Arabia.* Paris: OECD Publishing. https://doi.org/10.1787/76df15a2-en.

———. 2021. *TALIS: Positive, High-Achieving Students? What Schools and Teachers Can Do*. Paris: OECD Publishing. https://doi.org/10.1787/3b9551db-en.

Page, Elspeth, and Jyotsna Jha, eds. 2009. *Exploring the Bias: Gender and Stereotyping in Secondary Schools*. London: Commonwealth Secretariat.

Patrinos, Harry A., and Noam Angrist. 2019. "Harmonized Learning Outcomes: Transforming Learning Assessment Data into National Education Policy Reforms." *World Bank Blogs.* August 12. https://blogs.worldbank.org/opendata/harmonized-learning-outcomes-transforming-learning-assessment-data-national-education.

Reynolds, David, Pam Sammons, Bieke De Fraine, Jan Van Damme, Tony Townsend, Charles Teddlie, and Sam Stringfield. 2014. "Educational Effectiveness Research (EER): A State-of-the-Art Review." *School Effectiveness and School Improvement* 25 (2): 197–230. https://doi.org/10.1080/09243453.2014.885450.

Ridge, Natasha, and Soohyun Jeon. 2020. "Father Involvement and Education in the Middle East: Geography, Gender, and Generations." *Comparative Education Review* 64 (4): 725–48. https://doi.org/10.1086/710768.

Rogers, Halsey F., and Emiliana Vegas. 2009. "No More Cutting Class? Reducing Teacher Absence and Providing Incentives for Performance." Policy Research Working Paper 4847, World Bank, Washington, DC. https://doi.org/10.1596/1813-9450-4847.

Rutkowski, Leslie, Eugenio Gonzalez, Marc Joncas, and Matthias von Davier. 2010. "International Large-Scale Assessment Data: Issues in Secondary Analysis and Reporting." *Educational Researcher* 39 (2): 142–51. https://doi.org/10.3102/0013189X10363170.

Stromquist, Nelly. 2007. *The Gender Socialization Process in Schools: A Cross-National Comparison (2008/ED/EFA/MRT/PI/71)*. Background paper prepared for the Education for All Global Monitoring Report 2008, Education for All by 2015: Will we make it?

Teixeira, Pedro J., Marta M. Marques, Marlene N. Silva, Jennifer Brunet, Joan L. Duda, Leen Haerens, Jennifer La Guardia, et al. 2020. "A Classification of Motivation and Behavior Change Techniques Used in Self-Determination Theory-Based Interventions in Health Contexts." *Motivation Science* 6 (4): 438–55. https://doi.org/10.1037/mot0000172.

World Bank. 2012. *World Development Report 2012: Gender Equality and Development*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/4391.

———. 2019. *Ending Learning Poverty: What Will It Take?* Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/32553.

———. 2021. EdStats: Education Statistics (database). Washington, DC: World Bank. https://datatopics.worldbank.org/education/.

Younger, Mike, and Mary Cobbett. 2014. "Gendered Perceptions of Schooling: Classroom Dynamics and Inequalities within Four Caribbean Secondary Schools." *Educational Review* 66 (1): 1–21. https://doi.org/10.1080/00131911.2012.749218.

Appendix A1 – Supplementary data from NALO 2018

Table A1.1 Student reports of selected variables, by school gender

			Male (%)	Female (%)	φ/φ_c
Grade 4	Repeated a year at school because of poor results		14.5	6.6	.129***
	How many books have you read?	*0*	22.6	12.0	.148***
		1-5	27.8	33.3	
		6-10	15.7	14.4	
		>10	33.8	40.4	
	Math is important in life	*A lot*	81.4	84.1	.051***
		Somewhat	13.3	12.6	
		Never	5.3	3.3	
Grade 8	Repeated a year at school because of poor results		6.9	4.5	.051***
	How many books have you read?	*0*	30.6	19.3	.137***
		1-5	42.6	53.1	
		6-10	10.6	11.6	
		>10	16.2	16.1	
	Math is important in life	*A lot*	47.6	36.3	.115***
		Somewhat	38.6	48.0	
		Never	13.8	15.7	

Note: *p < .05; **p < .01; ***p < .001.

Table A1.2. Percentage of students *not born in Saudi Arabia* attending schools of various types (parent reports), by school gender

		Male (%)	Female (%)	φ/φ_c
Grade 4	Government school	85.0	95.2	.206***
	Quran school	1.1	2.2	
	Private school	13.9	2.6	
Grade 8	Government school	87.7	96.3	.168***
	Quran school	1.3	1.0	
	Private school	11.1	2.7	

Note: *p < .05; **p < .01; ***p < .001.

Table A1.3. Percentage of teachers 'always' agreeing with selected statements, by school gender

		Male (%)	Female (%)	φ/φ_c
Grade 4	Parents know the [academic] level of the student	32.8	62.8	.302***
	Students respect their [academically] excellent classmates	38.2	47.0	.094***
	Students are keen to [academically] excel in school	26.5	31.6	.071***
Grade 8	Parents know the [academic] level of the student	22.6	40.3	.236***
	Students respect their [academically] excellent classmates	34.8	44.0	.115***
	Students are keen to [academically] excel in school	21.5	16.5	.065***

Note: *p < .05; **p < .01; ***p < .001.

Table A1.4 School leaders' reports of selected variables, by school gender

		Male (%)	Female (%)	φ/φ_c
Grade 4	There is a school library from which students can borrow	67.3	48.5	.224***
	Parents' expectations of students' performance have been achieved *(agree or strongly agree)*	63.6	74.3	.178***
	Parental support to improve students' performance *(high or very high)*	44.1	51.5	.155***
Grade 8	There is a school library from which students can borrow	58.5	62.1	.059***
	Parents' expectations of students' performance have been achieved *(agree or strongly agree)*	49.1	59.8	.183***
	Parental support to improve students' performance *(high or very high)*	40.2	50.7	.158***

Note: *p < .05; **p < .01; ***p < .001.

Appendix A2

Table A2.1. TIMSS cut-scores for categories of continuous indices, TIMSS 2019

	Index	lowest category	cut-score	middle category	cut-score	highest category
Grade 4	Home resources for learning	few resources	7.4	some resources	11.8	many resources
	Literacy and numeracy readiness for school	not ready	8.6	moderately ready	11.2	very ready
	Student likes learning mathematics	do not like	8.4	somewhat like	10.2	very much like
	Student likes learning science		7.6		9.7	
	Student confident in mathematics	not confident	8.5	somewhat confident	10.7	very confident
	Student confident in science		8.2		10.2	
	Student sense of school belonging	little sense of school belonging	7.2	some sense of school belonging	9.6	high sense of school belonging
	Bullying	about weekly	7.4	about monthly	9.2	never or almost never
	Safe and orderly schools	less than safe and orderly	6.8	somewhat safe and orderly	9.9	very safe and orderly
	School emphasis on academic success	medium emphasis	9.2	high emphasis	13.0	very high emphasis
	School discipline	moderate to severe problems	7.6	minor problems	9.7	hardly any problems
	Teacher job satisfaction	less than satisfied	6.5	somewhat satisfied	10.1	very satisfied
Grade 8	Home educational resources	few resources	8.4	some resources	12.2	many resources
	Student likes learning mathematics	do not like	9.4	somewhat like	11.4	very much like
	Student likes learning science		8.3		10.6	
	Student confident in mathematics	not confident	9.5	somewhat confident	12.1	very confident
	Student confident in science		9.2		11.3	
	Student sense of school belonging	little sense of school belonging	7.8	some sense of school belonging	10.7	high sense of school belonging
	Bullying	about weekly	7.2	about monthly	8.8	never or almost never
	Safe and orderly schools	less than safe and orderly	7.3	somewhat safe and orderly	10.5	very safe and orderly
	School emphasis on academic success	medium emphasis	9.6	high emphasis	13.1	very high emphasis
	School discipline	moderate to severe problems	8.0	minor problems	10.8	hardly any problems
	Teacher job satisfaction	less than satisfied	6.8	somewhat satisfied	10.2	very satisfied

Appendix A3

Table A3.1. Hierarchical two-level linear regression model for mathematics achievement, grade 4, TIMSS 2019

		Step 1		Step 2		Step 3		Step 4		Step 5		Step 6		Step 7	
R^2	student-level (%)	1.3		6.7		19.9		20.3		20.7		21.0		26.2	
	class/school-level (%)			12.9		9.7		13.7		37.7		54.9		71.4	
Intercept (*SE*)		392.48 (6.83)		406.84 (7.59)		424.25 (7.58)		430.25 (8.21)		425.93 (17.05)		396.29 (21.00)		296.27 (37.86)	
Student-level variables (reference category)		***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***
Student gender (male)		19.73 (8.67)*	0.23 (0.10)*	7.68 (7.95)	0.09 (0.09)	-7.42 (7.98)	-0.09 (0.10)	-17.55 (8.99)	-0.21 (0.11)	-16.21 (10.50)	-0.19 (0.12)	-6.71 (11.92)	-0.08 (0.14)	164.98 (54.64)**	1.87 (0.59)**
Student immigration status (native)															
second-generation immigrant students				3.42 (8.36)	0.04 (0.10)	4.91 (8.28)	0.06 (0.10)	6.19 (8.55)	0.07 (0.10)	11.95 (10.66)	0.14 (0.13)	15.40 (11.36)	0.18 (0.13)	14.58 (11.44)	0.17 (0.13)
first-generation immigrant students				48.05 (12.55)***	0.57 (0.15)***	30.19 (13.04)*	0.36 (0.16)*	28.08 (13.44)*	0.33 (0.16)*	30.49 (15.55)	0.36 (0.18)	26.74 (17.41)	0.31 (0.20)	26.53 (17.24)	0.30 (0.20)
Home resources for learning				3.67 (1.59)*	0.06 (0.03)*	2.19 (1.68)	0.04 (0.03)	2.53 (1.90)	0.04 (0.03)	4.09 (1.95)*	0.07 (0.03)*	4.35 (2.11)*	0.07 (0.04)*	4.64 (2.11)*	0.08 (0.03)*
Student owns mobile phone (no)				-6.36 (3.65)	-0.08 (0.04)	-5.79 (3.71)	-0.07 (0.04)	-3.27 (4.02)	-0.04 (0.05)	-1.55 (4.33)	-0.02 (0.05)	-0.10 (4.70)	0.00 (0.06)	0.10 (4.51)	0.00 (0.05)
Literacy and numeracy readiness for school				8.54 (1.15)***	0.20 (0.03)***	6.44 (1.13)***	0.15 (0.03)***	6.37 (1.10)***	0.15 (0.03)***	7.60 (1.26)***	0.18 (0.03)***	7.57 (1.44)***	0.17 (0.03)***	7.57 (1.43)***	0.17 (0.03)***
Preschool attendance and duration (did not attend)															
1 year or less				0.63 (4.41)	0.01 (0.05)	1.55 (4.31)	0.02 (0.05)	3.22 (4.84)	0.04 (0.06)	-1.45 (6.21)	-0.02 (0.07)	1.51 (6.97)	0.02 (0.08)	0.38 (6.81)	0.00 (0.08)
2 years				0.98 (5.92)	0.01 (0.07)	0.94 (5.61)	0.01 (0.07)	-0.52 (6.26)	-0.01 (0.07)	-7.03 (8.65)	-0.08 (0.10)	-6.20 (8.84)	-0.07 (0.10)	-8.60 (8.84)	-0.10 (0.10)
3 years or more				7.94 (7.40)	0.09 (0.09)	12.75 (7.10)	0.15 (0.08)	17.52 (7.39)*	0.21 (0.09)*	11.22 (8.52)	0.13 (0.10)	15.10 (9.06)	0.18 (0.11)	14.00 (8.81)	0.16 (0.10)
Student absenteeism (never or almost never)															
once every two months						6.95 (6.27)	0.08 (0.08)	5.58 (6.57)	0.07 (0.08)	-1.03 (9.02)	-0.01 (0.11)	5.36 (9.85)	0.06 (0.12)	4.74 (9.69)	0.05 (0.11)
once a month						-1.82 (5.32)	-0.02 (0.06)	0.45 (5.91)	0.01 (0.07)	7.61 (7.90)	0.09 (0.09)	7.54 (8.60)	0.09 (0.10)	7.21 (8.49)	0.08 (0.10)
once every two weeks						-16.69 (7.09)*	-0.20 (0.08)*	-18.18 (7.34)*	-0.22 (0.09)*	-18.31 (8.51)*	-0.22 (0.10)*	-20.34 (8.88)*	-0.24 (0.10)*	-21.87 (8.85)*	-0.25 (0.10)*
once a week						-28.14 (5.28)***	-0.33 (0.06)***	-26.40 (5.41)***	-0.31 (0.06)***	-23.86 (6.20)***	-0.28 (0.07)***	-23.53 (7.35)**	-0.28 (0.08)**	-34.61 (8.67)***	-0.39 (0.10)***

Student-level variables (reference category)	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***
Student likes learning mathematics					4.24 (1.19)***	0.10 (0.03)***	4.29 (1.24)**	0.10 (0.03)**	5.05 (1.53)**	0.12 (0.04)**	4.67 (1.60)**	0.11 (0.04)**	4.82 (1.55)**	0.11 (0.04)**
Student confident in mathematics					9.93 (0.99)***	0.26 (0.03)***	9.26 (1.10)***	0.24 (0.03)***	7.83 (1.38)***	0.20 (0.04)***	7.91 (1.50)***	0.20 (0.04)***	8.24 (1.47)***	0.21 (0.04)***
Student sense of school belonging							-2.81 (0.93)**	-0.08 (0.03)**	-3.36 (1.11)**	-0.09 (0.03)**	-3.10 (1.16)**	-0.08 (0.03)**	-3.10 (1.17)**	-0.08 (0.03)**
Bullying[a]							4.59 (0.97)***	0.12 (0.03)***	4.94 (1.21)***	0.13 (0.03)***	4.91 (1.28)***	0.13 (0.04)***	5.20 (1.26)***	0.14 (0.03)***
Class/school-level variables (reference category)														
School mean of home resources for learning			17.17 (7.57)*	0.28 (0.13)*	15.00 (7.12)*	0.26 (0.13)*	13.99 (7.41)	0.24 (0.13)	11.57 (7.23)	0.22 (0.14)	15.70 (9.03)	0.30 (0.17)	10.65 (8.07)	0.20 (0.15)
School location (urban)														
suburban/medium size city or large town			-15.48 (9.42)	-0.31 (0.18)	-9.66 (9.34)	-0.21 (0.20)	-10.95 (9.30)	-0.24 (0.20)	-7.71 (9.82)	-0.18 (0.22)	-8.40 (11.43)	-0.19 (0.26)	-3.81 (9.56)	-0.09 (0.21)
small town or village/remote rural			30.03 (16.86)	0.60 (0.32)	24.71 (15.79)	0.53 (0.32)	18.96 (16.94)	0.41 (0.35)	24.20 (18.82)	0.55 (0.41)	54.91 (26.49)*	1.23 (0.53)*	42.47 (24.77)	0.94 (0.53)
Teacher age (40 years or older)														
29 years or younger			-23.61 (13.60)	-0.47 (0.27)	-20.04 (14.40)	-0.43 (0.31)	-21.58 (15.25)	-0.47 (0.33)	9.73 (20.67)	0.22 (0.47)	15.59 (19.55)	0.35 (0.44)	-4.73 (18.68)	-0.11 (0.42)
30–39 years			0.05 (8.43)	0.00 (0.17)	0.32 (8.02)	0.01 (0.17)	0.74 (8.31)	0.02 (0.18)	13.18 (10.96)	0.30 (0.25)	16.37 (14.17)	0.37 (0.32)	18.35 (12.26)	0.41 (0.27)
Safe and orderly schools							3.04 (2.22)	0.12 (0.09)	2.43 (3.32)	0.11 (0.15)	0.32 (3.21)	0.01 (0.14)	6.63 (4.00)	0.28 (0.16)
School emphasis on academic success							1.53 (2.11)	0.07 (0.10)	-0.18 (2.58)	-0.01 (0.13)	-0.26 (2.76)	-0.01 (0.14)	2.93 (2.58)	0.15 (0.13)
School discipline							0.73 (1.86)	0.04 (0.09)	-3.26 (3.58)	-0.18 (0.20)	-3.91 (3.80)	-0.22 (0.21)	-1.60 (3.74)	-0.09 (0.20)
Teacher years of experience									1.25 (0.59)*	0.23 (0.11)*	1.76 (0.60)**	0.33 (0.11)**	2.23 (0.59)***	0.41 (0.11)***
Teacher job satisfaction									1.53 (4.16)	0.04 (0.12)	1.10 (4.34)	0.03 (0.12)	-0.03 (4.03)	0.00 (0.11)
Time assigned to mathematics homework (15 minutes or less)[b]									12.58 (8.62)	0.29 (0.20)	4.63 (8.85)	0.10 (0.20)	9.68 (8.52)	0.21 (0.19)

Class/school-level variables (reference category)	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*
Poor teacher timekeeping (not a problem)														
minor problem									-25.79 (11.57)*	-0.59 (0.27)*	-24.39 (11.49)*	-0.55 (0.27)*	-14.89 (8.67)	-0.33 (0.19)
moderate or serious problem									-9.48 (14.95)	-0.22 (0.34)	-2.18 (17.35)	-0.05 (0.39)	13.44 (16.06)	0.30 (0.36)
Teacher absenteeism (not a problem)														
minor problem									2.67 (13.69)	0.06 (0.31)	8.04 (13.45)	0.18 (0.31)	5.03 (11.02)	0.11 (0.24)
moderate or serious problem									-6.17 (23.06)	-0.14 (0.53)	-7.64 (23.90)	-0.17 (0.54)	-8.53 (21.01)	-0.19 (0.47)
Teacher major area of study (education and mathematics)														
mathematics but not education									-8.33 (9.88)	-0.19 (0.23)	-8.20 (9.26)	-0.18 (0.21)	-6.76 (8.18)	-0.15 (0.18)
all other majors									-23.35 (13.66)	-0.54 (0.31)	-9.88 (12.49)	-0.22 (0.28)	-18.21 (10.74)	-0.40 (0.23)
Professional development hours on mathematics (more than 35 hours)														
16–35 hours									-24.59 (11.01)*	-0.56 (0.26)*	-29.46 (11.69)*	-0.66 (0.26)*	-22.62 (10.50)*	-0.50 (0.23)*
6–15 hours									3.28 (12.46)	0.08 (0.28)	-5.35 (13.24)	-0.12 (0.30)	4.99 (12.26)	0.11 (0.27)
less than 6 hours									-13.12 (17.06)	-0.30 (0.39)	-3.30 (14.44)	-0.08 (0.33)	4.14 (12.28)	0.09 (0.27)
none									6.89 (16.99)	0.16 (0.39)	-1.71 (21.24)	-0.04 (0.48)	-2.60 (16.21)	-0.06 (0.36)
Professional development on mathematics content (no)									14.92 (11.55)	0.34 (0.25)	19.53 (10.85)	0.44 (0.23)	15.17 (10.60)	0.33 (0.23)
Professional development on mathematics pedagogy (no)									7.51 (11.18)	0.17 (0.26)	13.29 (11.26)	0.30 (0.25)	14.85 (9.46)	0.33 (0.22)

Class/school-level variables (reference category)	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*
Principal years of experience											-0.13 (0.79)	-0.02 (0.12)	-0.29 (0.71)	-0.04 (0.10)
Principal highest level of education (ISCED levels 7 & 8 — master's or doctorate degree)[c]											-3.76 (15.01)	-0.08 (0.34)	4.78 (12.47)	0.11 (0.28)
Principal qualification in educational leadership (no)											-4.76 (13.02)	-0.11 (0.29)	-6.06 (10.83)	-0.13 (0.24)
School library (no)											19.09 (10.82)	0.43 (0.23)	15.10 (10.29)	0.33 (0.22)
Interaction terms (reference category)														
Student gender*Safe and orderly schools													-16.25 (5.13)**	-1.13 (0.34)**
Student gender*Student absenteeism (never or almost never) — once a week													22.88 (10.44)*	0.26 (0.12)*
Student gender*Teacher age (40 years or older) — 29 years or younger													89.46 (30.62)**	1.01 (0.34)**
Fit statistics	Loglikelihood (H_0)	-32207.65		-18375.32		-17089.57		-14821.53		-9654.01	-8549.92		-8533.11	
	AIC	64423.29		36784.64		34225.13		29699.05		19394.02	17193.84		17166.23	
	BIC	64449.71		36887.46		34362.86		29862.73		19627.00	17442.83		17431.12	

Note: Null model: *Intercept* (*SE*): 402.02 (4.44), H_0 = -32210.89, *AIC* = 64427.78, *BIC* = 64447.59. *p < .05; **p < .01; ***p < .001. [a]Higher scores indicate less frequent bullying; [b]Other category: 16 minutes or more; [c]Other category: ISCED Level 6 — bachelor's or equivalent level.

Table A3.2. Hierarchical two-level linear regression model for science achievement, grade 4, TIMSS 2019

		Step 1		Step 2		Step 3		Step 4		Step 5		Step 6		Step 7	
R^2	student-level (%)	7.6		12.2		21.0		21.4		22.7		22.5		27.4	
	class/school-level (%)			16.3		17.4		20.3		36.6		46.0		64.8	
Intercept (*SE*)		381.71 (6.89)		397.44 (8.37)		415.31 (7.92)		426.93 (8.46)		427.69 (18.15)		433.35 (21.20)		320.48 (35.66)	
Student-level variables (reference category)		***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***
Student gender (male)		53.14 (8.29)***	0.55 (0.08)***	39.36 (7.04)***	0.42 (0.07)***	21.61 (6.57)**	0.23 (0.07)**	17.44 (7.38)*	0.19 (0.08)*	16.31 (8.29)*	0.18 (0.09)*	21.52 (8.79)*	0.23 (0.09)*	249.50 (59.64)***	2.59 (0.59)***
Student immigration status (native)															
second-generation immigrant students				12.11 (8.52)	0.13 (0.09)	14.70 (9.16)	0.16 (0.10)	16.13 (10.18)	0.18 (0.11)	11.93 (11.74)	0.13 (0.13)	14.12 (11.98)	0.15 (0.13)	13.25 (12.08)	0.14 (0.13)
first-generation immigrant students				44.37 (9.32)***	0.47 (0.10)***	36.26 (10.55)**	0.39 (0.11)**	30.06 (12.10)*	0.33 (0.13)*	37.85 (13.30)**	0.41 (0.14)**	32.10 (14.51)*	0.34 (0.16)*	37.67 (13.70)**	0.39 (0.14)**
Home resources for learning				4.19 (1.57)**	0.06 (0.02)**	3.11 (1.78)	0.05 (0.03)	3.31 (1.88)	0.05 (0.03)	2.58 (2.43)	0.04 (0.04)	2.85 (2.79)	0.04 (0.04)	2.83 (2.79)	0.04 (0.04)
Student owns mobile phone (no)				-3.96 (1.11)***	-0.04 (0.01)***	-3.16 (1.24)*	-0.03 (0.01)*	-2.30 (1.69)	-0.03 (0.02)	-1.99 (1.94)	-0.02 (0.02)	-1.58 (1.96)	-0.02 (0.02)	-1.45 (1.96)	-0.02 (0.02)
Literacy and numeracy readiness for school				9.51 (1.17)***	0.20 (0.02)***	7.29 (1.07)***	0.16 (0.02)***	7.11 (1.10)***	0.15 (0.02)***	7.33 (1.32)***	0.16 (0.03)***	7.62 (1.41)***	0.16 (0.03)***	10.35 (1.91)***	0.21 (0.04)***
Preschool attendance and duration (did not attend)															
1 year or less				-0.81 (5.24)	-0.01 (0.06)	1.64 (4.92)	0.02 (0.05)	1.83 (4.87)	0.02 (0.05)	5.25 (5.08)	0.06 (0.05)	4.88 (5.19)	0.05 (0.06)	3.79 (5.14)	0.04 (0.05)
2 years				10.23 (6.02)	0.11 (0.06)	10.44 (6.19)	0.11 (0.07)	8.92 (6.66)	0.10 (0.07)	11.80 (6.88)	0.13 (0.07)	8.61 (6.40)	0.09 (0.07)	6.55 (6.36)	0.07 (0.07)
3 years or more				2.03 (7.38)	0.02 (0.08)	7.78 (7.02)	0.08 (0.07)	6.36 (7.75)	0.07 (0.08)	10.07 (8.27)	0.11 (0.09)	8.71 (8.62)	0.09 (0.09)	8.66 (8.49)	0.09 (0.09)
Student absenteeism (never or almost never)															
once every two months						8.03 (5.90)	0.09 (0.06)	4.09 (6.50)	0.04 (0.07)	4.43 (7.50)	0.05 (0.08)	7.90 (7.20)	0.09 (0.08)	8.14 (7.23)	0.08 (0.08)
once a month						2.63 (5.09)	0.03 (0.05)	-1.46 (6.42)	-0.02 (0.07)	-2.88 (7.24)	-0.03 (0.08)	-2.91 (7.42)	-0.03 (0.08)	-2.24 (7.39)	-0.02 (0.08)
once every two weeks						-16.74 (7.30)*	-0.18 (0.08)*	-18.50 (7.85)*	-0.20 (0.09)*	-16.35 (8.85)	-0.18 (0.10)	-15.47 (9.37)	-0.17 (0.10)	-15.07 (9.20)	-0.16 (0.10)
once a week						-22.77 (4.75)***	-0.24 (0.05)***	-22.91 (4.79)***	-0.25 (0.05)***	-22.69 (5.41)***	-0.24 (0.06)***	-21.05 (5.67)***	-0.23 (0.06)***	-20.78 (5.50)***	-0.22 (0.06)***
Student likes learning science						4.95 (1.27)***	0.12 (0.03)***	4.86 (1.34)***	0.12 (0.03)***	5.43 (1.66)**	0.14 (0.04)**	5.33 (1.76)**	0.13 (0.04)**	5.16 (1.75)**	0.13 (0.04)**

Student-level variables (reference category)	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***
Student confident in science					8.98 (1.23)***	0.20 (0.03)***	7.85 (1.24)***	0.18 (0.03)***	8.16 (1.52)***	0.18 (0.03)***	7.89 (1.62)***	0.17 (0.04)***	8.09 (1.61)***	0.17 (0.04)***
Student sense of school belonging							-3.58 (1.10)**	-0.09 (0.03)**	-3.88 (1.27)**	-0.09 (0.03)**	-3.68 (1.32)**	-0.09 (0.03)**	-3.55 (1.31)**	-0.08 (0.03)**
Bullying[a]							5.71 (1.24)***	0.14 (0.03)***	6.17 (1.36)***	0.15 (0.03)***	5.78 (1.44)***	0.14 (0.03)***	5.91 (1.44)***	0.14 (0.03)***
Class/school-level variables (reference category)														
School mean of home resources for learning			21.38 (6.67)**	0.36 (0.11)**	22.12 (6.12)***	0.41 (0.11)***	18.81 (6.50)**	0.35 (0.12)**	20.20 (6.97)**	0.38 (0.13)**	22.16 (6.74)**	0.43 (0.13)**	17.82 (5.67)**	0.33 (0.11)**
School location (urban)														
suburban/medium size city or large town			-8.85 (8.22)	-0.18 (0.17)	-5.86 (7.26)	-0.13 (0.16)	-8.89 (8.17)	-0.20 (0.18)	-15.52 (9.10)	-0.34 (0.20)	-10.96 (9.66)	-0.25 (0.21)	-36.03 (10.25)***	-0.79 (0.20)***
small town or village/remote rural			37.58 (14.65)*	0.77 (0.28)*	31.37 (13.32)*	0.71 (0.28)*	26.80 (14.68)	0.61 (0.32)	35.84 (17.51)*	0.80 (0.37)*	60.77 (20.36)**	1.37 (0.44)**	60.10 (16.90)***	1.32 (0.36)***
Teacher age (40 years or older)														
29 years or younger			-29.23 (13.04)*	-0.60 (0.26)*	-25.94 (12.86)*	-0.59 (0.28)*	-30.06 (14.19)*	-0.68 (0.31)*	-48.67 (19.04)*	-1.08 (0.40)*	-43.78 (17.99)*	-0.98 (0.39)*	-17.31 (16.16)	-0.38 (0.36)
30–39 years			-8.08 (7.94)	0.17 (0.16)	-9.41 (7.37)	-0.21 (0.16)	-13.84 (8.02)	-0.31 (0.17)	-26.42 (11.01)*	-0.59 (0.24)*	-28.52 (10.26)**	-0.64 (0.23)**	-20.29 (9.24)*	-0.45 (0.20)*
Safe and orderly schools							1.61 (2.34)	0.07 (0.11)	4.41 (2.74)	0.20 (0.12)	5.29 (3.07)	0.24 (0.13)	9.92 (3.52)**	0.44 (0.14)**
School emphasis on academic success							1.38 (1.83)	0.07 (0.09)	-1.17 (1.95)	-0.05 (0.09)	-1.91 (1.83)	-0.09 (0.09)	-0.75 (1.71)	-0.04 (0.08)
School discipline							0.22 (1.60)	0.01 (0.08)	-1.67 (2.55)	-0.09 (0.13)	-1.84 (2.68)	-0.10 (0.14)	-0.85 (2.51)	-0.04 (0.13)
Teacher years of experience									-1.18 (0.67)	-0.22 (0.12)	-1.51 (0.60)*	-0.28 (0.11)*	-0.95 (0.59)	-0.17 (0.11)
Teacher job satisfaction									-2.64 (3.25)	-0.08 (0.09)	-3.32 (3.34)	-0.10 (0.10)	-4.52 (3.17)	-0.13 (0.09)
Time assigned to science homework (15 minutes or less)[b]									-3.54 (10.40)	-0.08 (0.23)	-7.30 (11.31)	-0.17 (0.25)	-5.47 (9.81)	-0.12 (0.22)

Class/school-level variables (reference category)	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*
Poor teacher timekeeping (not a problem)														
minor problem									3.63 (10.27)	0.08 (0.23)	2.50 (11.75)	0.06 (0.26)	4.86 (10.50)	0.11 (0.23)
moderate or serious problem									1.00 (16.37)	0.02 (0.36)	-1.66 (17.19)	-0.04 (0.39)	-0.39 (15.45)	-0.01 (0.34)
Teacher absenteeism (not a problem)														
minor problem									-7.15 (11.19)	-0.16 (0.25)	1.01 (13.17)	0.02 (0.30)	-2.16 (10.94)	-0.05 (0.24)
moderate or serious problem									-16.14 (14.69)	-0.36 (0.32)	1.87 (16.55)	0.04 (0.37)	5.55 (13.59)	0.12 (0.30)
Teacher major area of study (education and science)														
science but not education									-2.38 (9.82)	-0.05 (0.22)	-2.69 (9.78)	-0.06 (0.22)	-7.35 (8.84)	-0.16 (0.19)
all other majors									-16.82 (17.36)	-0.37 (0.38)	-23.80 (18.47)	-0.53 (0.41)	-34.13 (17.72)	-0.75 (0.38)
Professional development hours on science (more than 35 hours)														
16–35 hours									5.55 (14.21)	0.12 (0.31)	3.50 (14.09)	0.08 (0.32)	5.25 (11.73)	0.11 (0.26)
6–15 hours									2.88 (11.07)	0.06 (0.24)	-5.01 (12.01)	-0.11 (0.27)	-4.52 (10.92)	-0.10 (0.24)
less than 6 hours									13.39 (15.77)	0.30 (0.34)	18.81 (15.70)	0.42 (0.35)	12.82 (13.31)	0.28 (0.29)
none									-13.09 (17.23)	-0.29 (0.39)	-35.32 (15.06)*	-0.79 (0.35)*	-37.77 (15.23)*	-0.83 (0.35)*
Professional development on science content (no)									11.97 (11.22)	0.27 (0.24)	9.96 (12.76)	0.22 (0.28)	11.64 (11.14)	0.26 (0.24)
Professional development on science pedagogy (no)									6.66 (9.66)	0.15 (0.21)	-2.53 (10.00)	-0.06 (0.22)	-1.60 (9.00)	-0.04 (0.20)
Principal years of experience											-0.20 (0.66)	-0.03 (0.10)	0.15 (0.60)	0.02 (0.09)

Class/school-level variables (reference category)	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*
Principal highest level of education (ISCED levels 7 & 8 — master's or doctorate degree)[c]											-11.91 (10.93)	-0.27 (0.25)	-13.19 (9.41)	-0.29 (0.21)
Principal qualification in educational leadership (no)											5.72 (9.43)	0.13 (0.21)	7.28 (8.66)	0.16 (0.19)
School library (no)											4.69 (8.59)	0.11 (0.19)	7.65 (7.90)	0.17 (0.17)
Interaction terms (reference category)														
Student gender*Safe and orderly schools													-15.61 (4.93)**	-0.94 (0.29)**
Student gender*Literacy and numeracy readiness for school													-6.29 (2.57)*	-0.35 (0.15)*
Student gender*School location (urban) — suburban/medium size city or large town													75.76 (15.57)***	0.79 (0.16)***
Fit statistics	Loglikelihood (H_0)	-32681.95		-24789.30		-22223.42		-17688.83		-13820.91	-12381.76		-12361.03	
	AIC	65371.91		49612.60		44492.84		35433.66		27727.81	24857.51		24822.05	
	BIC	65398.32		49720.30		44636.27		35601.96		27975.68	25123.27		25104.77	

Note: Null model: *Intercept* (*SE*): 407.42 (4.88), H_0 = -32704.43, *AIC* = 65414.86, *BIC* = 65434.67. *p < .05; **p < .01; ***p < .001. [a]Higher scores indicate less frequent bullying; [b]Other category: 16 minutes or more; [c]Other category: ISCED Level 6 — bachelor's or equivalent level.

Table A3.3. Hierarchical two-level regression model for mathematics achievement, grade 8, TIMSS 2019

		Step 1		Step 2		Step 3		Step 4		Step 5		Step 6		Step 7	
R^2	student-level (%)	2.3		6.2		25.3		25.3		25.2		26.7		27.3	
	class/school-level (%)			39.5		40.7		45.4		63.4		64.0		74.4	
Intercept (*SE*)		384.54 (4.27)		393.90 (6.09)		408.61 (5.96)		410.32 (6.24)		406.71 (16.40)		367.09 (23.42)		310.38 (36.78)	
Student-level variables (reference category)		***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***
Student gender (male)		20.78 (5.29)***	0.30 (0.08)***	15.79 (4.78)**	0.23 (0.07)**	18.14 (4.52)***	0.26 (0.07)***	15.47 (4.92)**	0.22 (0.07)**	9.88 (6.06)	0.14 (0.09)	16.29 (6.98)*	0.23 (0.10)*	90.11 (35.49)*	1.27 (0.49)*
Student immigration status (native)															
second-generation immigrant students				27.44 (6.04)***	0.40 (0.09)***	22.17 (5.17)***	0.32 (0.07)***	22.45 (5.52)***	0.32 (0.08)***	20.24 (7.56)**	0.29 (0.11)**	17.59 (7.92)*	0.25 (0.11)*	17.51 (7.97)*	0.25 (0.11)*
first-generation immigrant students				55.40 (6.87)***	0.80 (0.10)***	42.56 (6.82)***	0.61 (0.10)***	43.43 (7.11)***	0.63 (0.10)***	34.47 (8.79)***	0.49 (0.12)***	31.83 (9.37)**	0.45 (0.13)**	31.92 (9.33)**	0.45 (0.13)**
Home educational resources				6.68 (0.92)***	0.15 (0.02)***	3.85 (0.83)***	0.08 (0.02)***	3.65 (0.93)***	0.08 (0.02)***	4.17 (1.12)***	0.09 (0.02)***	4.09 (1.03)***	0.09 (0.02)***	4.10 (1.03)***	0.09 (0.02)***
Student owns mobile phone (no)				-3.47 (3.52)	-0.05 (0.05)	0.03 (3.74)	0.00 (0.05)	0.10 (3.77)	0.00 (0.05)	-1.32 (4.82)	-0.02 (0.07)	4.20 (5.13)	0.06 (0.07)	3.80 (5.07)	0.05 (0.07)
Student absenteeism (never or almost never)															
once every two months						-5.25 (4.70)	-0.08 (0.07)	-6.73 (4.44)	-0.10 (0.06)	-7.08 (5.85)	-0.10 (0.08)	-8.73 (5.81)	-0.12 (0.08)	-9.10 (5.78)	-0.13 (0.08)
once a month						-16.71 (4.18)***	-0.24 (0.06)***	-17.72 (3.96)***	-0.26 (0.06)***	-17.28 (5.18)**	-0.25 (0.07)**	-19.47 (5.06)***	-0.28 (0.07)***	-20.02 (5.20)***	-0.28 (0.07)***
once every two weeks						-20.82 (3.74)***	-0.30 (0.05)***	-22.48 (3.86)***	-0.33 (0.06)***	-27.00 (4.68)***	-0.38 (0.07)***	-29.92 (4.84)***	-0.42 (0.07)***	-30.31 (4.83)***	-0.43 (0.07)***
once a week						-39.72 (4.98)***	-0.57 (0.07)***	-41.01 (5.15)***	-0.59 (0.08)***	-38.91 (6.36)***	-0.55 (0.09)***	-44.23 (6.40)***	-0.63 (0.09)***	-44.52 (6.53)***	-0.63 (0.09)***
Student likes learning mathematics						-4.76 (1.07)***	-0.14 (0.03)***	-4.25 (1.12)***	-0.13 (0.03)***	-4.15 (1.49)**	-0.12 (0.04)**	-4.30 (1.57)**	-0.13 (0.05)**	-4.35 (1.57)**	-0.13 (0.05)**
Student confident in mathematics						15.27 (0.94)***	0.44 (0.03)***	15.04 (1.02)***	0.43 (0.03)***	15.27 (1.13)***	0.44 (0.03)***	15.24 (1.27)***	0.44 (0.03)***	15.26 (1.27)***	0.44 (0.03)***
Student sense of school belonging								-1.62 (0.79)*	-0.05 (0.02)*	-1.22 (1.04)	-0.03 (0.03)	-1.03 (1.12)	-0.03 (0.03)	-1.00 (1.12)	-0.03 (0.03)
Bullying[a]								0.86 (0.57)	0.03 (0.02)	-0.37 (0.85)	-0.01 (0.03)	-0.42 (0.91)	-0.01 (0.03)	-0.43 (0.91)	-0.01 (0.03)
Time spent on mathematics homework (15 minutes or less)[b]										-11.19 (3.70)**	-0.16 (0.05)**	-11.35 (3.98)**	-0.16 (0.06)**	-11.40 (4.01)**	-0.16 (0.06)**

Class/school-level variables (reference category)	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*
School mean of home educational resources			27.21 (3.70)***	0.61 (0.08)***	24.39 (3.40)***	0.60 (0.08)***	22.21 (3.67)***	0.52 (0.08)***	23.74 (4.24)***	0.57 (0.10)***	25.97 (5.32)***	0.62 (0.11)***	26.91 (5.29)***	0.57 (0.11)***
School location (urban)														
suburban/medium size city or large town			-6.63 (5.37)	-0.21 (0.17)	-8.79 (4.86)	-0.30 (0.16)	-10.80 (5.03)*	-0.35 (0.16)*	3.47 (5.28)	0.11 (0.17)	6.64 (6.19)	0.22 (0.20)	9.63 (6.22)	0.28 (0.17)
small town or village/remote rural			0.83 (7.39)	0.03 (0.23)	-2.18 (6.84)	-0.07 (0.23)	-4.79 (7.36)	-0.16 (0.24)	-0.98 (7.13)	-0.03 (0.24)	2.05 (11.06)	0.06 (0.36)	4.58 (11.13)	0.13 (0.32)
Teacher age (40 years or older)														
29 years or younger			-9.81 (8.36)	-0.30 (0.25)	-11.77 (7.88)	-0.39 (0.26)	-12.72 (8.30)	-0.41 (0.27)	-22.96 (15.40)	-0.76 (0.49)	-20.08 (19.13)	-0.66 (0.62)	-23.36 (19.05)	-0.68 (0.54)
30–39 years			-3.95 (4.75)	-0.12 (0.15)	-4.00 (4.50)	-0.13 (0.15)	-0.61 (4.85)	-0.02 (0.16)	-6.02 (9.72)	-0.20 (0.32)	-5.54 (11.15)	-0.18 (0.36)	-6.51 (11.35)	-0.19 (0.33)
Safe and orderly schools							0.57 (1.03)	0.04 (0.07)	1.85 (1.38)	0.12 (0.08)	0.87 (1.68)	0.05 (0.10)	0.87 (1.62)	0.05 (0.09)
School emphasis on academic success							2.92 (1.11)**	0.20 (0.07)**	2.40 (1.57)	0.16 (0.10)	1.80 (1.70)	0.13 (0.12)	5.76 (2.48)*	0.36 (0.13)*
School discipline							0.05 (0.83)	0.00 (0.07)	-2.23 (1.50)	-0.21 (0.14)	-2.60 (1.92)	-0.24 (0.18)	-3.11 (1.93)	-0.25 (0.15)
Teacher years of experience									-0.28 (0.69)	-0.07 (0.16)	-0.49 (0.88)	-0.12 (0.21)	-0.60 (0.88)	-0.13 (0.19)
Teacher job satisfaction									-1.87 (2.21)	-0.08 (0.09)	-1.07 (2.18)	-0.05 (0.09)	-1.70 (2.15)	-0.07 (0.08)
Time assigned to mathematics homework (15 minutes or less)[b]									6.50 (5.12)	0.22 (0.17)	10.39 (6.03)	0.34 (0.20)	13.77 (6.21)*	0.40 (0.17)*
Poor teacher timekeeping (not a problem)														
minor problem									6.14 (7.12)	0.21 (0.23)	7.38 (8.34)	0.25 (0.27)	7.72 (7.64)	0.23 (0.22)
moderate or serious problem									-1.58 (11.47)	-0.05 (0.38)	1.67 (15.21)	0.06 (0.50)	10.83 (14.07)	0.32 (0.40)
Teacher absenteeism (not a problem)														
minor problem									-2.58 (6.76)	-0.09 (0.22)	-3.46 (7.99)	-0.12 (0.26)	-1.30 (7.43)	-0.04 (0.22)
moderate or serious problem									-20.51 (11.34)	-0.68 (0.37)	-20.00 (13.91)	-0.66 (0.45)	-28.83 (14.24)*	-0.84 (0.39)*

Class/school-level variables (reference category)	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***
Teacher highest level of education (Up to ISCED level 6 — bachelor's or equivalent level)[c]									28.40 (13.35)*	0.95 (0.46)*	54.32 (15.71)**	1.80 (0.56)**	67.81 (17.26)***	1.99 (0.52)***
Teacher major area of study (mathematics and mathematics education)														
mathematics but not mathematics education									-2.03 (6.44)	-0.07 (0.21)	-5.55 (7.31)	-0.18 (0.23)	-5.02 (7.25)	-0.15 (0.21)
all other majors									4.00 (8.74)	0.14 (0.30)	-0.44 (9.16)	-0.01 (0.30)	-5.21 (9.65)	-0.15 (0.27)
Professional development hours on mathematics (more than 35 hours)														
16–35 hours									-15.52 (7.44)*	-0.52 (0.24)*	-14.62 (8.62)	-0.48 (0.27)	-18.01 (8.67)*	-0.53 (0.24)*
6–15 hours									-4.68 (7.23)	-0.15 (0.24)	-0.91 (7.97)	-0.03 (0.26)	-2.98 (7.67)	-0.08 (0.22)
less than 6 hours									-7.32 (8.94)	-0.25 (0.30)	-9.99 (8.94)	-0.33 (0.29)	-11.23 (8.69)	-0.33 (0.25)
none									-22.28 (11.69)	-0.75 (0.39)	-19.37 (11.70)	-0.64 (0.39)	-24.04 (12.16)*	-0.71 (0.35)*
Professional development on mathematics content (no)									7.65 (4.86)	0.25 (0.16)	4.33 (5.31)	0.14 (0.17)	5.13 (5.06)	0.15 (0.15)
Professional development on mathematics pedagogy (no)									-4.76 (6.15)	-0.16 (0.21)	-0.02 (7.05)	-0.01 (0.23)	1.28 (6.95)	0.03 (0.20)
Principal years of experience											-0.24 (0.38)	-0.06 (0.09)	-0.34 (0.39)	-0.07 (0.08)
Principal highest level of education (ISCED levels 7 & 8 — master's or doctorate degree)[d]											-4.32 (11.75)	-0.15 (0.39)	1.36 (11.86)	0.03 (0.35)
Principal qualification in educational leadership (no)											12.85 (8.06)	0.42 (0.27)	11.58 (7.68)	0.34 (0.22)
School library (no)											9.32 (6.98)	0.31 (0.22)	9.76 (6.95)	0.28 (0.19)

Interaction terms (reference category)		*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*
Student gender*School emphasis on academic success														-6.43 (3.04)*	-0.56 (0.26)*
Fit statistics	Loglikelihood (H_0)	-32290.17		-27131.49		-25465.23		-23557.67		-13670.53		-12113.38		-12109.54	
	AIC	64588.35		54288.99		50968.47		47163.34		27423.06		24316.77		24311.07	
	BIC	64614.93		54373.18		51090.67		47315.85		27661.20		24572.77		24572.76	

Note: Null model: *Intercept* (*SE*): 394.83 (2.88), H_0 = -32297.78, *AIC* = 64601.56, *BIC* = 64621.49. *p < .05; **p < .01; ***p < .001. [a]Higher scores indicate less frequent bullying; [b]Other category: 16 minutes or more; [c]Other category: ISCED levels 7 & 8 — master's or doctorate degree; [d]Other category: ISCED Level 6 — bachelor's or equivalent level.

Table A3.4. Hierarchical two-level regression model for science achievement, grade 8, TIMSS 2019

	Step 1		Step 2		Step 3		Step 4		Step 5		Step 6		Step 7	
R^2 student-level (%)	9.6		11.8		23.1		23.0		24.0		25.5		27.2	
R^2 class/school-level (%)			37.9		39.1		45.6		65.5		69.8		79.0	
Intercept (*SE*)	406.87 (4.78)		412.26 (5.99)		427.80 (6.01)		437.21 (6.16)		462.45 (13.44)		441.38 (19.49)		426.88 (24.75)	
Student-level variables (reference category)	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***
Student gender (male)	50.40 (5.80)***	0.62 (0.07)***	45.83 (5.26)***	0.57 (0.06)***	42.28 (5.00)***	0.53 (0.06)***	35.42 (5.03)***	0.45 (0.06)***	34.88 (7.10)***	0.44 (0.08)***	39.89 (7.52)***	0.50 (0.09)***	74.88 (25.07)**	0.92 (0.31)**
Student immigration status (native)														
second-generation immigrant students			25.22 (6.31)***	0.31 (0.08)***	22.15 (5.85)***	0.28 (0.07)***	20.70 (6.13)**	0.26 (0.08)**	33.19 (7.73)***	0.42 (0.10)***	31.81 (8.36)***	0.40 (0.11)***	31.84 (8.36)***	0.39 (0.11)***
first-generation immigrant students			53.93 (7.72)***	0.67 (0.09)***	42.57 (7.72)***	0.53 (0.10)***	44.95 (8.34)***	0.57 (0.10)***	44.76 (8.96)***	0.56 (0.11)***	43.75 (9.67)***	0.55 (0.12)***	43.76 (9.57)***	0.54 (0.12)***
Home educational resources			7.03 (0.93)***	0.13 (0.02)***	4.17 (0.89)***	0.08 (0.02)***	4.16 (0.93)***	0.08 (0.02)***	5.16 (1.34)***	0.10 (0.03)***	5.39 (1.40)***	0.10 (0.03)***	5.44 (1.40)***	0.10 (0.03)***
Student owns mobile phone (no)			-0.25 (3.97)	0.00 (0.05)	-0.04 (4.05)	0.00 (0.05)	-1.35 (4.12)	-0.02 (0.05)	-3.04 (5.90)	-0.04 (0.07)	-0.32 (6.32)	0.00 (0.08)	-1.32 (6.53)	-0.02 (0.08)
Student absenteeism (never or almost never)														
once every two months					2.46 (4.59)	0.03 (0.06)	2.54 (4.77)	0.03 (0.06)	-3.73 (6.04)	-0.05 (0.08)	-5.55 (6.26)	-0.07 (0.08)	-5.16 (6.26)	-0.06 (0.08)
once a month					-7.15 (4.32)	-0.09 (0.05)	-7.93 (4.29)	-0.10 (0.05)	-11.55 (5.26)*	-0.15 (0.07)*	-13.34 (5.34)*	-0.17 (0.07)*	-12.49 (5.24)*	-0.15 (0.07)*
once every two weeks					-10.48 (5.31)*	-0.13 (0.07)*	-12.70 (5.83)*	-0.16 (0.07)*	-9.67 (7.23)	-0.12 (0.09)	-10.40 (7.47)	-0.13 (0.09)	-10.19 (7.49)	-0.13 (0.09)
once a week					-37.93 (4.58)***	-0.47 (0.06)***	-39.38 (4.68)***	-0.50 (0.06)***	-44.29 (6.50)***	-0.56 (0.08)***	-46.77 (6.44)***	-0.58 (0.08)***	-47.10 (6.46)***	-0.58 (0.08)***
Student likes learning science					-0.38 (0.86)	-0.01 (0.02)	-0.02 (0.91)	0.00 (0.03)	-0.57 (1.11)	-0.02 (0.03)	-0.85 (1.17)	-0.02 (0.03)	-0.79 (1.18)	-0.02 (0.03)
Student confident in science					11.66 (0.98)***	0.29 (0.02)***	10.98 (1.08)***	0.28 (0.03)***	11.14 (1.39)***	0.28 (0.03)***	11.35 (1.54)***	0.28 (0.04)***	14.09 (2.02)***	0.35 (0.05)***
Student sense of school belonging							-2.94 (0.85)**	-0.07 (0.02)**	-1.09 (1.29)	-0.03 (0.03)	-1.20 (1.45)	-0.03 (0.04)	-1.15 (1.45)	-0.03 (0.04)
Bullying[a]							3.73 (0.75)***	0.10 (0.02)***	1.77 (1.19)	0.05 (0.03)	1.49 (1.21)	0.04 (0.03)	1.32 (1.21)	0.03 (0.03)
Time spent on science homework (15 minutes or less)[b]									-11.92 (3.65)**	-0.15 (0.05)**	-12.37 (3.99)**	-0.15 (0.05)**	-11.94 (3.99)**	-0.15 (0.05)**

Class/school-level variables (reference category)	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*
School mean of home educational resources			28.03 (4.16)***	0.61 (0.08)***	25.49 (4.12)***	0.60 (0.09)***	21.87 (4.54)***	0.51 (0.10)***	21.33 (4.13)***	0.55 (0.10)***	20.23 (4.66)***	0.55 (0.11)***	18.59 (4.37)***	0.46 (0.11)***
School location (urban)														
suburban/medium size city or large town			-7.02 (5.65)	-0.21 (0.17)	-7.48 (5.15)	-0.25 (0.17)	-8.50 (5.19)	-0.28 (0.17)	-7.82 (6.08)	-0.27 (0.21)	-6.53 (5.96)	-0.24 (0.22)	-7.84 (5.70)	-0.26 (0.19)
small town or village/remote rural			0.35 (8.28)	0.01 (0.25)	-2.71 (8.22)	-0.09 (0.27)	-2.24 (8.52)	-0.07 (0.27)	0.61 (8.14)	0.02 (0.28)	3.12 (8.44)	0.11 (0.31)	-3.79 (8.53)	-0.13 (0.28)
Teacher age (40 years or older)														
29 years or younger			-9.55 (10.34)	-0.29 (0.31)	-8.28 (9.17)	-0.27 (0.30)	-17.24 (9.12)	-0.56 (0.30)	1.49 (11.87)	0.05 (0.41)	9.78 (13.38)	0.35 (0.48)	3.56 (12.88)	0.12 (0.43)
30–39 years			-1.06 (4.80)	-0.03 (0.14)	-1.13 (4.59)	-0.04 (0.15)	-3.97 (4.72)	-0.13 (0.15)	-0.81 (6.27)	-0.03 (0.22)	-2.14 (6.47)	-0.08 (0.23)	-20.34 (7.79)**	-0.68 (0.23)**
Safe and orderly schools							2.45 (1.24)*	0.16 (0.08)*	3.01 (1.31)*	0.20 (0.08)*	2.29 (1.40)	0.16 (0.09)	1.96 (1.44)	0.13 (0.09)
School emphasis on academic success							2.50 (1.13)*	0.17 (0.08)*	0.45 (1.35)	0.03 (0.10)	0.07 (1.24)	0.01 (0.10)	0.46 (1.29)	0.03 (0.09)
School discipline							0.13 (0.96)	0.01 (0.08)	0.07 (1.16)	0.01 (0.11)	0.64 (1.14)	0.07 (0.12)	0.13 (1.13)	0.01 (0.11)
Teacher years of experience									0.56 (0.47)	0.15 (0.12)	0.54 (0.48)	0.15 (0.13)	0.54 (0.51)	0.14 (0.13)
Teacher job satisfaction									-3.47 (2.41)	-0.15 (0.10)	-3.19 (2.47)	-0.14 (0.11)	-3.35 (2.36)	-0.14 (0.10)
Time assigned to science homework (15 minutes or less)[b]									-8.64 (6.37)	-0.30 (0.22)	-4.98 (6.32)	-0.18 (0.23)	-7.03 (5.91)	-0.23 (0.19)
Poor teacher timekeeping (not a problem)														
minor problem									-0.02 (8.55)	0.00 (0.30)	2.15 (8.32)	0.08 (0.30)	0.82 (7.67)	0.03 (0.26)
moderate or serious problem									8.14 (11.38)	0.28 (0.39)	13.22 (11.46)	0.48 (0.41)	12.78 (10.64)	0.42 (0.36)
Teacher absenteeism (not a problem)														
minor problem									-3.51 (7.01)	-0.12 (0.24)	1.19 (6.79)	0.04 (0.25)	2.17 (6.24)	0.07 (0.21)
moderate or serious problem									-21.43 (8.61)*	-0.75 (0.29)*	-17.19 (8.26)*	-0.62 (0.29)*	-16.62 (7.64)*	-0.55 (0.25)*

Class/school-level variables (reference category)	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***	***B (SE)***	***β (SE)***
Teacher highest level of education (Up to ISCED level 6 — bachelor's or equivalent level)[c]									-5.26 (12.99)	-0.18 (0.45)	-3.91 (13.17)	-0.14 (0.48)	-8.79 (12.29)	-0.29 (0.40)
Teacher major area of study (science and science education)														
science but not science education									-10.46 (6.76)	-0.37 (0.24)	-11.49 (6.62)	-0.42 (0.24)	-11.88 (6.43)	-0.40 (0.22)
all other majors									12.21 (11.37)	0.43 (0.39)	22.96 (10.73)*	0.83 (0.37)*	20.62 (9.38)*	0.69 (0.31)*
Professional development hours on science (more than 35 hours)														
16–35 hours									0.68 (7.57)	0.02 (0.26)	-3.53 (7.63)	-0.13 (0.28)	-3.66 (7.06)	-0.12 (0.24)
6–15 hours									3.92 (6.53)	0.14 (0.22)	-0.71 (6.72)	-0.03 (0.24)	-4.61 (6.57)	-0.15 (0.22)
less than 6 hours									0.83 (8.63)	0.03 (0.30)	-0.04 (8.80)	0.00 (0.32)	3.27 (8.49)	0.11 (0.28)
none									0.73 (14.42)	0.03 (0.50)	3.65 (15.01)	0.13 (0.54)	6.03 (14.52)	0.20 (0.48)
Professional development on science content (no)									7.01 (7.01)	0.25 (0.25)	8.62 (7.78)	0.32 (0.28)	11.07 (7.34)	0.37 (0.24)
Professional development on science pedagogy (no)									-3.52 (8.32)	-0.12 (0.29)	-0.02 (8.96)	0.00 (0.33)	0.27 (7.98)	0.01 (0.27)
Principal years of experience											-0.09 (0.35)	-0.02 (0.09)	-0.19 (0.35)	-0.05 (0.08)
Principal highest level of education (ISCED levels 7 & 8 — master's or doctorate degree)[d]											4.73 (13.27)	0.17 (0.48)	6.23 (13.36)	0.21 (0.45)
Principal qualification in educational leadership (no)											3.02 (6.33)	0.11 (0.23)	-0.05 (6.46)	0.00 (0.22)
School library (no)											6.66 (6.20)	0.24 (0.23)	9.22 (5.81)	0.31 (0.20)

Interaction terms (reference category)		*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*	*B (SE)*	*β (SE)*
Student gender*Student confident in science														-4.88 (1.91)*	-0.35 (0.13)*
Student gender*Teacher age (40 years or older) — 30-39 years														32.15 (10.20)**	0.40 (0.12)**
Fit statistics	Loglikelihood (H_0)	-32968.21		-29618.39		-27632.53		-25284.64		-14301.28		-13016.61		-13004.99	
	AIC	65944.43		59262.78		55303.05		50617.28		28684.55		26123.23		26103.99	
	BIC	65971.01		59262.78		55426.29		50770.87		28923.63		26381.44		26373.68	

Note: Null model: *Intercept* (*SE*): 431.76 (3.69), H_0 = -33003.98, *AIC* = 66013.96, *BIC* = 66033.89. *p < .05; **p < .01; ***p < .001. [a]Higher scores indicate less frequent bullying; [b]Other category: 16 minutes or more; [c]Other category: ISCED levels 7 & 8 — master's or doctorate degree; [d]Other category: ISCED Level 6 — bachelor's or equivalent level.