

RESEARCH

Open Access



# Boys' underachievement in mathematics and science: An analysis of national and international assessment data from the Kingdom of Saudi Arabia

Mahmoud A. A. Elsayed<sup>1\*</sup>, Aidan Clerkin<sup>2</sup> , Vasiliki Pitsia<sup>2,3</sup> , Nayyaf Aljabri<sup>4,5</sup> and Khaleel Al-Harbi<sup>5</sup>

\*Correspondence:  
[melsayed@worldbank.org](mailto:melsayed@worldbank.org)

<sup>1</sup> The World Bank, Washington, DC, USA

<sup>2</sup> Educational Research Centre, Dublin, Ireland

<sup>3</sup> Department of Education, School of Education, University of Nicosia, Nicosia, Cyprus

<sup>4</sup> Department of Educational Administration, College of Education, Taibah University, Medina, Kingdom of Saudi Arabia

<sup>5</sup> Education and Training Evaluation Commission, Riyadh, Saudi Arabia

## Abstract

Boys in the Kingdom of Saudi Arabia consistently and significantly underperform compared to girls across different grades and subjects, forming one of the largest gender gaps in student achievement in the world. Saudi Arabia offers a unique setting in which boys and girls attend separate schools on a universal basis starting from grade 1. This means that boys and girls are educated only by male and female teachers, respectively, in effect inhabiting parallel education systems. In this context, this study examines the factors that are associated with student achievement in mathematics and science in grades 4 and 8 and the extent to which these associations are different for boys and girls, in an effort to gain insights into boys' underachievement in mathematics and science in Saudi Arabia. The paper employs data from two recent large-scale assessments of education: Trends in International Mathematics and Science Study (TIMSS) 2019 and Saudi Arabia's National Assessment of Learning Outcomes (NALO) 2018. A series of hierarchical two-level linear regression models showed 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 indicated 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, significant interactions between gender and the extent to which students feel confident 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.

**Keywords:** Gender, Boys' underachievement, Mathematics, Science, TIMSS, NALO, Hierarchical two-level linear regression modelling, Saudi Arabia

## Background

The Kingdom of Saudi Arabia 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 underperformed girls by approximately 17 points in mathematics and 47 points in science. Furthermore, in grade 4, 53% of boys failed to achieve minimum proficiency in mathematics, compared to 45% of girls. Similarly, in grade 8, 58% of boys did not achieve minimal proficiency in mathematics, compared to 49% of girls. Data from another assessment, the Programme for International Student Assessment (PISA) 2018, show a similar finding that almost twice as many 15-year-old boys (65%) as girls (38%) failed to achieve minimum proficiency in reading (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—sometimes in favor of boys, and sometimes in favor of girls (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 & Pan, 2013; Buchmann & DiPrete, 2006; DiPrete & Buchmann, 2013; DiPrete & Jennings, 2012; Fortin et al., 2015; Jha & Pouezevara, 2016; Legewie & 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 & DiPrete, 2006; Cornwell et al., 2013; DiPrete & Jennings, 2012; Downey & Vogt Yuan, 2005; OECD, 2021). Other studies have found that school characteristics such as school quality, disciplinary practices, and school climate—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 (defined as school-level gains in mathematics and reading scores) on the gender achievement gap between opposite-gender siblings who attend the same sets of schools (Autor et al., 2016). This study shows 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 shows 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 demonstrates that increases in teachers' perceptions of classroom discipline problems were associated with an increase in the achievement gender gap (OECD, 2021). Other school organizational issues, such as poor learning conditions and organizational problems, are also found to exacerbate the gender gap.

Such findings indicate that boys' achievement tends to be negatively impacted by challenging learning conditions to a greater extent than girls' achievement, and are

consistent with previous work on gender gaps in attitudes showing that girls, in general, tend to report more positive attitudes toward learning (DiPrete & Buchmann, 2013). Previous research also shows that social norms, gender stereotypes, and teacher and school expectations contribute to the gender gap in performance (Jha & Pouezevara, 2016; Page & Jha, 2009; Stromquist, 2007; Younger & Cobbett, 2014). In other words, teachers and schools contribute to developing and reinforcing different expectations of appropriate behavior for boys and girls, which, in turn, may hinder boys' performance.

Another group of studies have investigated the relationship between the gender composition of schools and/or classes and students' outcomes. The findings from this research show weak associations between single-sex schooling and academic achievement, with some exceptions in certain grades and among certain subpopulations of students. Pahlke et al. (2014), for example, synthesized available literature on the effects of single-sex compared with co-educational schooling on a wide range of student outcomes, such as mathematics and science achievement, gender stereotypes, interpersonal relations, and students' perceptions. After reviewing more than 400 studies, the authors conclude that "there is little evidence of an advantage of SS [single-sex] schooling for girls or boys for any of the outcomes" (p. 1064). Similarly, using data from the Secondary Entrance Assessment and the Caribbean Secondary Education Certification (CSEC) examination in Trinidad and Tobago, Jackson (2012) shows that students attending single-sex secondary schools perform no better than those attending co-educational secondary schools, except for female students with strong preferences for single-sex education who perform better on the CSEC examination (at grade 10). Pahlke et al. (2013), using the random assignment of students into single-sex and co-educational schools in Korea to study the effect of single-sex education on student achievement in mathematics and science, have found no association between school gender and student achievement. Other studies, however, show some positive effects of single-sex education, especially among females. Following an approach similar to Pahlke et al. (2013), Eisenkopf et al. (2015) used the random assignment of female students into single-sex and co-educational secondary schools in Switzerland to examine the effect of single-sex education on students' academic performance. Their results suggest that single-sex education improves females' performance in mathematics and their self-confidence. Booth and Nolen (2012) have also found that girls in single-sex schools are likely to be more competitive than girls attending co-educational schools, but the same relationship has not been found for boys.

### **The current study**

Building on the existing research, this study examines (i) the contribution of a range of student, family, class, and school variables in predicting overall Saudi student achievement in mathematics and science at grades 4 (primary school) and 8 (intermediate/middle school), and (ii) the extent to which these variables contribute to the large observed gender gap in Saudi student performance in these subjects at grades 4 and 8. Data from both TIMSS 2019 and Saudi Arabia's National Assessment of Learning Outcomes (NALO) 2018 were employed. Both TIMSS and NALO assess fourth- and eighth-grade students' mathematics and science achievement, as well as collecting a wide range of contextual information about students, their families, teachers, and schools.

This study contributes to the existing 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 Middle East and North Africa (MENA) region, and especially in Saudi Arabia. In an effort to provide this evidence, this study uses 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. It, therefore, provides important insights for policy makers, both in Saudi Arabia and in other countries in the MENA region, regarding factors that may contribute to the observed gender gaps in mathematics and science achievement.

Secondly, Saudi Arabia offers a unique setting in which boys and girls attend separate schools on a universal basis starting from grade 1. This means that boys and girls are educated only by male and female teachers, respectively, meaning that, in effect, they inhabit parallel education systems. This is displayed in Fig. 1, which shows that Saudi Arabia is the only country among all countries participating in TIMSS 2019 to have a completely gender-segregated education system.<sup>1</sup> 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, therefore, 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.

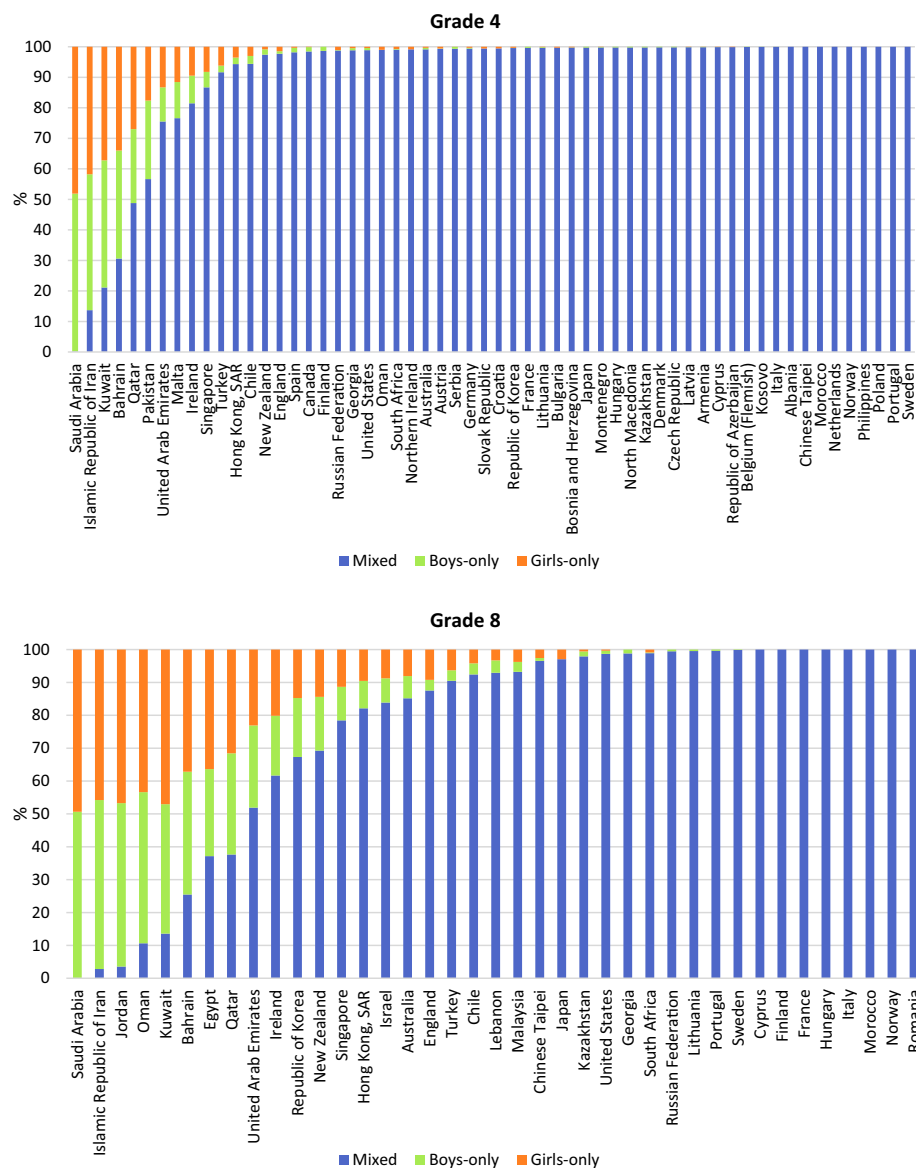
While this paper exploits this feature of the Saudi education system in its analysis, it also acknowledges potential difficulties in interpreting findings due to this extreme degree of separation, as gender differences signify differences between schools attended only by boys and schools only attended by girls, and not differences among individual students. Additionally, 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 in an environment that is completely separate from the training and work environment of the (female) teachers of girls.

### **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 onwards and are taught

---

<sup>1</sup> 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 that this number will increase in the coming years.



**Fig. 1** Distribution of students in single-gender or mixed education among countries participating in TIMSS 2019

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 may be 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 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 Ministry 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 – the total enrolment in primary school expressed as a percentage of the total primary school-aged population – increased from 58% in 1979 to 101% in 2019. During the same period, the GER in secondary education increased from 27 to 112%.<sup>2</sup>

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 53<sup>rd</sup> of 58 countries in grade 4 and 37<sup>th</sup> of 39 countries in grade 8. In PISA 2018, less than half (48%) of 15-year-old students in Saudi Arabia achieved minimum proficiency in reading and almost no student was a top performer (i.e., achieving proficiency levels 5 or 6; OECD, 2019). Additionally, only 27% of Saudi students in the same age group achieved at least minimum proficiency in mathematics, compared to an Organization for Economic Co-operation and Development (OECD) average of 75%. Notably, learning outcomes in Saudi Arabia are low relative to the country's level of wealth, and Saudi Arabia has been identified as an outlier when examining Harmonized Learning Outcomes<sup>3</sup> relative to gross domestic product (GDP) worldwide (Patrinos & Angrist, 2019). Finally, as noted above, the gender differences in achievement observed in Saudi Arabia are consistently among the largest in the world, with girls showing a consistent advantage over boys across all grade levels and subject areas.

## Methods

### 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 and high-quality nationally representative sample, while NALO was conducted to provide regionally representative information within Saudi Arabia (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.<sup>4</sup>

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 this paper is conducted using TIMSS 2019 data. Given the high degree of overlap in 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

<sup>2</sup> GER can exceed 100% as it may include students who are younger or older than the official age cohort for primary or secondary school.

<sup>3</sup> 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).

<sup>4</sup> 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.



in TIMSS. In this way, NALO 2018 data are used as supplementary information to shed additional light on questions arising from the multilevel analysis of TIMSS data.

### ***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. 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% male) across 220 public and private schools and 5,680 grade 8 students (mean age 13.9 years; 49.2% male) across 209 schools took part in 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 et al., 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 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<sup>5</sup> for mathematics and science scores to ensure accurate population-level estimates of achievement and facilitate appropriate analyses taking the complex nature of the data into account.

### ***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. Following a similar approach to TIMSS, NALO also collects contextual information through student, parent, teacher, and school questionnaires.

In NALO 2018, 27,985 grade 4 students (50.2% male) across 964 government, private, and Quran schools completed tests of mathematics and science, as did 30,157 grade 8 students (49.6% male) across 939 schools. The schools that took part in NALO were sampled using procedures similar to those used in TIMSS. Also following the procedures in

<sup>5</sup> 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 student 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.

TIMSS, sampling weights are provided by ETEC to ensure that the NALO database is weighted to represent the national population, and plausible values and replicate weights are provided for appropriate calculation of achievement data.

### **Measures and variables**

In this study, data from the TIMSS mathematics and science tests as well as the TIMSS student, parent, teacher, and school questionnaires were used for the primary analysis, and data from the NALO mathematics and science tests and the student, teacher, and school questionnaires were used to complement the primary analysis.

### **Outcome variables**

Students' mathematics and science achievement constituted the outcome variables. In TIMSS, the scores for each student across the two subjects and grades are reported on scales with international centrepoin ts set at 500 and standard deviations (*SD*) at 100, with most scores falling within the 300–700 band. Following a similar scaling approach to TIMSS, NALO scores for each student across the two subjects and grades are reported on scales with national averages set at 500 and *SD* at 100.

### **Predictor variables**

A range of predictor variables related to student, teacher, and school demographics and home background, student engagement and attitudes, school climate, teacher qualifications and practices, and school leadership and resources were included in the analysis. Information about these variables can be found in the Additional file 1: Table S1.

### **Statistical analysis**

Prior to the main analysis, descriptive statistics were computed, and a series of 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 ( $\phi$ ) and the Cramer's *V* ( $\phi_c$ ) effect size measures were used for the contextual categorical variables for  $2 \times 2$  contingency tables and for contingency tables larger than  $2 \times 2$ , respectively. The Cohen's *d* effect size measure was used for the contextual continuous variables (Fritz et al., 2012). Cohen's (1988) guidelines were used in conjunction with Hattie's (2009) guidelines for the interpretation of effect sizes. The IEA International Database Analyzer (IDB Analyzer) (IEA, 2021) and the Statistical Package for the Social Sciences (SPSS) (IBM Corporation, 2020) were used to compute estimates for TIMSS and NALO, respectively.

Next, hierarchical two-level linear regression analysis was conducted to explore the different factors that are associated with all students' achievement in mathematics and science in grades 4 and 8 more generally as well as boys' underperformance in Saudi Arabia. The regression models draw on the TIMSS 2019 data. Four models were constructed: (i) grade 4 mathematics, (ii) grade 4 science, (iii) grade 8 mathematics, and (iv) grade 8 science. Students were the unit of analysis at level 1 and classes were the unit of analysis at level 2. However, as the number of intact classes sampled from each school varied, with some schools having one class and other schools having two classes



**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
Gender	✓	✓	✓	✓	✓	✓	✓
Demographics and home background		✓	✓	✓	✓	✓	✓
Student engagement and attitudes			✓	✓	✓	✓	✓
School climate				✓	✓	✓	✓
Teacher qualifications and practices					✓	✓	✓
School leadership and resources						✓	✓
Interaction terms							✓

✓ indicates variables included in each step

sampled, the class level is confounded with the school level; in other words, for schools where only one intact class is available, the class level is identical to the school level. For this reason, two-level rather than three-level analysis was conducted, with classes constituting the level-2 unit of the analysis. However, the possible confounding of class and school levels was taken into account in conducting the analysis (e.g., calculation of the sampling weights) and in the interpretation of the results.

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) and teacher and school principal questionnaires (level 2 of the analysis). Variables were selected for inclusion a priori, based on previous literature on student achievement and the gender gap in education in particular, as discussed earlier, in addition to the expected theoretical or policy relevance of these variables to the question of gender differences 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 (Table 1). The first step of each model included gender only.<sup>6</sup> By including the 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 for the examination of their contribution in predicting overall achievement and in explaining the difference in achievement between boys and girls. In the final step of each model, the statistical significance of the interactions between 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

<sup>6</sup> While gender information in TIMSS is collected at the student level, in the case of Saudi Arabia, where schools are gender-segregated, it also reflects the school gender type.

plots shown below 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.

Equation 1 represents the random intercept null multilevel models (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  $j$ ,  $\beta_0$  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 random intercept multilevel models, which were applied to the TIMSS data, including  $v$  number of predictor variables at the student ( $ij$ ) or the class/school level ( $j$ ), while controlling for the clustering of the data.

$$y_{ij} = \beta_0 + u_{0j} + \beta_1 x_{1ij/j} + \dots + \beta_v x_{vij/j} + e_{ij} \quad (2)$$

where  $y_{ij}$  is the outcome variable (e.g., mathematics achievement) of student  $i$  in class/school  $j$ ,  $\beta_0$  is the mean intercept,  $u_{0j}$  is the variation of class/school  $j$  from the mean intercept,  $\beta_1$  is the regression slope for the predictor variable  $x_1$  of student  $i$  in class/school  $j$  or class/school  $j$ ,  $\beta_v$  is the regression slope for the predictor variable  $x_v$  of student  $i$  in class/school  $j$  or class/school  $j$ , and  $e_{ij}$  is the student-level residual error term. Additional file 1: Table S3 presents detailed equations for each step of the models.

All five 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. Given the imputation methodology of plausible values used to scale achievement data in TIMSS, there were no missing values in the achievement-related variables (i.e., mathematics and science achievement). Missing value analysis of the contextual variables showed that the average proportion of missing values across the variables included in the analysis at both grade 4 and grade 8 was 6.2% and that data were not missing at random; given this relatively low proportion of missing values and that data were not missing at random, these were not imputed. 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*), the default estimator for multilevel models with continuous outcomes in Mplus which provides estimates that are robust to the non-independence of observations (which, in this study, is a result of the clustering of students within classes/schools) (Muthén & Muthén 1998–2017). Multilevel linear regression analysis was performed using Mplus 8 (Muthén & Muthén 1998–2017).

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 ( $\beta$ s) 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.  $B_s$  are expressed in the original unit of each of the predictor variables, while  $\beta_s$  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.

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.

## Results

### Descriptive statistics

#### *Descriptive statistics for outcome variables*

Table 2 shows mean scores in mathematics and science by gender in the TIMSS 2019 and NALO 2018 assessments. Results from both datasets show that boys, or due to the gender-segregated school system in Saudi Arabia, schools attended by boys, consistently underperform compared to girls (or schools attended by 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.

#### *Descriptive statistics for 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 ( $m$ ), and  $SD$  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 ( $\phi/\phi_c$  and  $d$ ) for the gender differences in each of the contextual variables. As 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), thresholds and their corresponding interpretation along the continuum of each of these variables, as set by TIMSS, are provided in Additional file 1: Table S2.

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

		TIMSS 2019			NALO 2018		
		Male $m$ ( $SD$ )	Female $m$ ( $SD$ )	M-F	Male $m$ ( $SD$ )	Female $m$ ( $SD$ )	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

TIMSS 2019 international results in mathematics and science by Mullis et al. (2020). NALO data are authors' own calculations. All differences between genders have been rounded

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 having attended fewer hours of professional development on mathematics and science compared to teachers in girls' schools. In comparison to grade 4 girls, grade 4 boys report less positive attitudes (liking or feeling confident) toward mathematics and science, a lower sense of school belonging, and experience more frequent bullying. Additionally, boys' schools tend to be less safe and orderly, and teachers in boys' schools tend to report lower levels of job satisfaction.

At grade 8, both mathematics and science teachers are younger, on average, in boys' schools. 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 report attending 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.<sup>7</sup>

#### ***Additional contextual information for grades 4 and 8 – NALO 2018***

This section draws on data from NALO questions that were not available in TIMSS to enrich the description of students' experience of education in Saudi Arabia. All of the differences between boys and girls noted in this section are statistically significant at the 0.001 level. However, the effect sizes associated with these differences were generally small.

NALO data indicate that 14.5% of grade 4 boys had repeated a year at school because of poor academic performance, compared to 6.6% of grade 4 girls<sup>8</sup> (Table 5). In addition, grade 4 boys reported engaging in a lower level of reading than grade 4 girls. For example, 22.6% of boys reported never having read a book, compared to 12.0% of girls.

<sup>7</sup> The TIMSS variable describing teachers' qualifications for grade 4 shows that the 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.

<sup>8</sup> The relatively large proportion of grade 4 boys reporting having repeated a year at school because of poor academic performance is not consistent with existing data in Saudi Arabia. Provided that, and also that information on grade repetition is self-reported by students, which makes it subject to measurement error and other types of recall errors, especially at the primary level, these data should be interpreted with caution.

**Table 3** Contextual categorical variables by gender, TIMSS 2019

	Grade 4			Grade 8		
	Male	Female	$\phi/\phi_c$	Male	Female	$\phi/\phi_c$
	%	%		%	%	
<b>Student level variables</b>						
<i>Student immigration status</i>			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	
<i>Student owns mobile phone</i>			0.036**			0.000
Yes	62.7	58.3		81.4	82.3	
No	37.3	41.7		18.6	17.7	
<i>Preschool attendance and duration</i>			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				
<i>Student absenteeism</i>			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	
<i>Time spent on mathematics homework</i>						0.051**
15 min or less				75.8	73.3	
16 min or more				24.2	26.7	
<i>Time spent on science homework</i>						0.045**
15 min or less				71.0	77.4	
16 min or more				29.0	22.6	
<b>Class/school-level variables</b>						
<i>School location</i>			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	
<i>Teacher age (mathematics teacher)</i>			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	
<i>Teacher age (science teacher)</i>			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	
<i>Time assigned to mathematics homework</i>			0.147***			0.119***
15 min or less	52.7	67.2		58.4	62.8	
16 min or more	47.3	32.8		41.6	37.2	
<i>Time assigned to science homework</i>			0.054**			0.180***
15 min or less	77.3	69.8		66.8	81.4	
16 min or more	22.7	30.2		33.2	18.6	
<i>Poor teacher timekeeping</i>			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	

**Table 3** (continued)

	Grade 4			Grade 8		
	Male	Female	$\phi/\phi_c$	Male	Female	$\phi/\phi_c$
	%	%		%	%	
Moderate or serious problem	22.4	13.9		18.5	16.9	
<i>Teacher absenteeism</i>			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	
<i>Teacher highest level of education (mathematics teacher)</i>						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	
<i>Teacher highest level of education (science teacher)</i>						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	
<i>Teacher major area of study (mathematics teacher)</i>			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	
<i>Teacher major area of study (science teacher)</i>			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	
<i>Teacher major area of study (science teacher)</i>			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	
<i>Professional development hours on mathematics</i>			0.326***			0.196***
None	10.5	4.6		11.9	3.8	
Less than 6 h	20.1	5.4		14.2	12.2	
6–15 h	34.0	28.0		34.4	26.1	
16–35 h	19.2	27.1		20.9	31.3	
More than 35 h	16.2	34.9		18.5	26.7	
<i>Professional development hours on science</i>			0.238***			0.324***
None	18.6	8.7		13.5	0.2	
Less than 6 h	12.4	13.0		12.8	8.9	
6–15 h	32.9	28.1		35.2	26.0	
16–35 h	21.1	16.9		16.9	34.4	
More than 35 h	15.0	33.3		21.5	30.6	
<i>Professional development on mathematics content</i>			0.007			0.098***
Yes	55.3	55.0		42.0	50.0	



**Table 3** (continued)

	Grade 4			Grade 8		
	Male	Female	$\phi/\phi_c$	Male	Female	$\phi/\phi_c$
	%	%		%	%	
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	

\* $p < .05$

\*\* $p < .01$

\*\*\* $p < .001$

Conversely, 33.8% of boys had read more than 10 books, compared to 40.4% of girls. However, when asked whether mathematics is important in life, grade 4 boys and girls provided broadly similar levels of agreement (Table 5). 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%) than girls (2.6%) (Table 6). However, the vast majority of both boys (85.0%) and girls (95.2%) 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 were substantially less likely to agree that parents have a good understanding of their child's current academic level, suggesting a 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 ( $\phi = 0.302$ ; Table 7). Teachers in boys' schools are less likely to report that high-achieving students were respected among their peers compared to teachers in girls' schools (Table 7).

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 8). More grade 4 boys

**Table 4** Contextual continuous variables by gender, TIMSS 2019

	Grade 4				Grade 8			
	Male		Female		Male		Female	
	<i>m</i>	<i>SD</i>	<i>m</i>	<i>SD</i>	<i>m</i>	<i>SD</i>	<i>m</i>	<i>SD</i>
<b>Student-level variables</b>								
Home resources for learning (grade 4)/Home educational resources (grade 8)	9.3	1.41	9.5	1.34	9.4	1.59	9.5	1.51
Literacy and numeracy readiness for school	10.1	2.01	10.5	1.91				
Student likes learning mathematics	10.3	2.03	11.3	1.88	10.3	1.97	10.0	2.08
Student likes learning science	10.0	2.33	11.3	2.15	10.4	2.09	10.7	2.15
Student confident in mathematics	10.1	2.08	11.1	2.17	10.5	1.88	10.5	2.01
Student confident in science	10.0	2.01	11.0	2.05	10.5	1.92	10.9	2.02
Student sense of school belonging	10.0	2.33	11.1	2.01	10.2	2.09	10.3	1.92
Bullying	8.9	2.25	10.2	2.14	9.8	2.35	10.6	1.97
<b>Class/school-level variables</b>								
School mean of home resources for learning (grade 4)/School mean of home educational resources (grade 8)	9.2	0.83	9.5	0.68	9.4	0.69	9.5	0.67
Safe and orderly schools (mathematics teacher)	10.8	1.83	11.9	1.71	11.2	1.96	11.9	1.92
Safe and orderly schools (science teacher)	10.9	2.10	11.7	1.95	10.9	1.88	11.7	2.11
School emphasis on academic success	10.8	2.12	11.8	2.14	10.8	1.89	11.5	2.03
School discipline	9.5	2.40	10.6	2.15	10.0	2.68	11.6	2.37
Teacher years of experience (mathematics teacher)	15.9	8.43	14.4	7.70	10.6	7.29	12.2	7.27
Teacher years of experience (science teacher)	15.3	9.36	12.8	7.86	13.1	7.43	12.6	7.53
Teacher job satisfaction (mathematics teacher)	10.6	1.36	11.2	0.87	10.7	1.50	11.3	0.74
Teacher job satisfaction (science teacher)	10.6	1.50	11.3	0.88	10.5	1.65	11.5	0.64
Principal years of experience	9.9	6.65	8.6	7.09	9.4	7.56	8.6	8.30

\*  $p < .05$ \*\*  $p < .01$ \*\*\*  $p < .001$

**Table 5** Student reports of selected variables, by school gender, NALO 2018

		Male (%)	Female (%)	$\phi/\phi_c$
Grade 4	Repeated a year at school because of poor results	14.5	6.6	0.129
	How many books have you read?			0.148
	0	22.6	12.0	
	1–5	27.8	33.3	
	6–10	15.7	14.4	
	> 10	33.8	40.4	
	Math is important in life			0.051
	A lot	81.4	84.1	
Grade 8	Repeated a year at school because of poor results	6.9	4.5	0.051
	How many books have you read?			0.137
	0	30.6	19.3	
	1–5	42.6	53.1	
	6–10	10.6	11.6	
	> 10	16.2	16.1	
	Math is important in life			0.115
	A lot	47.6	36.3	
	Somewhat	38.6	48.0	
	Never	13.8	15.7	

All differences are statistically significant at the .001 level

**Table 6** Percentages of students not born in Saudi Arabia attending schools of various types (parent reports), by school gender, NALO 2018

		Male (%)	Female (%)	$\phi_c$
Grade 4	Government school	85.0	95.2	0.206
	Quran school	1.1	2.2	
	Private school	13.9	2.6	
Grade 8	Government school	87.7	96.3	0.168
	Quran school	1.3	1.0	
	Private school	11.1	2.7	

All differences are statistically significant at the .001 level

(67.3%) have access to a school library from which they can borrow compared to grade 4 girls (48.5%), despite girls themselves reporting reading more books, as noted above.

Boys in grade 8 were more likely (6.9%) than girls (4.5%) to report having repeated a year at school because of poor academic performance (Table 5). 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%) than girls (19.3%) 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% of boys agreeing a lot, compared to 36.3% of girls). Among immigrant students, grade 8 boys were much more likely to attend a private school (11.1%) than grade 8 girls (2.7%) (Table 6). Nonetheless, as at grade 4, most boys (87.7%) and girls (96.3%) 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

**Table 7** Percentages of teachers ‘always’ agreeing with selected statements, by school gender, NALO 2018

		Male (%)	Female (%)	$\phi$
Grade 4	Parents know the [academic] level of the student	32.8	62.8	0.302
	Students respect their [academically] excellent classmates	38.2	47.0	0.094
	Students are keen to [academically] excel in school	26.5	31.6	0.071
Grade 8	Parents know the [academic] level of the student	22.6	40.3	0.236
	Students respect their [academically] excellent classmates	34.8	44.0	0.115
	Students are keen to [academically] excel in school	21.5	16.5	0.065

All differences are statistically significant at the .001 level

**Table 8** School principals’ reports of selected variables, by school gender, NALO 2018

		Male (%)	Female (%)	$\phi$
Grade 4	There is a school library from which students can borrow	67.3	48.5	0.224
	Parents’ expectations of students’ performance have been achieved (agree or strongly agree)	63.6	74.3	0.178
	Parental support to improve students’ performance (high or very high)	44.1	51.5	0.155
Grade 8	There is a school library from which students can borrow	58.5	62.1	0.059
	Parents’ expectations of students’ performance have been achieved (agree or strongly agree)	49.1	59.8	0.183
	Parental support to improve students’ performance (high or very high)	40.2	50.7	0.158

All differences are statistically significant at the .001 level

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:  $\phi = 0.236$ ; Table 7). As at 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 8), and that boys’ parents’ expectations are being met to a lesser extent. There was little difference in grade 8 boys’ and girls’ access to a school library from which they can borrow books.

### Overall findings from multilevel models

Table 9 summarizes the main findings from the multilevel models. As discussed above, a series of hierarchical two-level linear regression models were constructed, starting with a simple model that includes gender but no other predictor variables. Then, with each step, the changes in the achievement gap between girls and boys, or due to the gender-segregated school system in Saudi Arabia, changes in the achievement gap between schools attended by girls and schools attended by boys, were 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 9 shows the coefficient for the gender difference. Regression estimates for the other predictor variables in steps 1–7 of each model are presented in

**Table 9** Summary of main findings from the hierarchical two-level linear regression models

	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
<b>Gender (female = 1)</b>						
Grade 4 mathematics	19.73* (8.67)	7.68 (7.95)	-7.42 (7.98)	-17.55 (8.99)	-16.21 (10.50)	-6.71 (11.92)
Grade 4 science	53.14 (8.29)***	39.36 (7.04)***	21.61 (6.57)**	17.44 (7.38)*	16.31 (8.29)*	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						✓

✓ indicates variables included in each step

Additional file 1: Tables S3–S7, while steps 6 and 7 are also presented and discussed in detail in the *Findings for grade 4* and *Findings for grade 8* sections below.

As shown in Table 9, boys underperformed relative 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 underperformed girls by 53 points in science compared to 20 points in mathematics. The results also show that, in grade 4, controlling for student, home, 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 are taken into account, the gap in mathematics achievement drops from 20 to 8 points and is no longer statistically significant (step 2).

However, in grade 8, controlling for a wide range of characteristics from the student, parent, teacher, and principal questionnaires explains a relatively small portion of the achievement gap. As shown in Table 9, 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 9 show that boys underperformed girls by 16 points in mathematics and 40 points in science, even after controlling for all observed characteristics.<sup>9</sup>

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 gender and each of the predictor variables was examined in step 7 of each model. The results from these interactions are presented in the final step (step 7) of each model, shown in Tables 10 and 11 below. 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

<sup>9</sup> The coefficients of the gender achievement gap in the grade 4 science, grade 8 mathematics, and grade 8 science models increase in step 6. When the interactions between gender and the variables within the *School leadership and resources* block were examined, none of them were statistically significant. Hence, this increase could not be attributed to an interaction between gender and these variables. A potential explanation for this increase is that the inclusion of additional variables in each of the models introduced new missing cases due to listwise deletion, which, in turn, may have had an impact on the gender coefficient.

orderly school climate is more strongly associated with improvements in boys' mathematics and science achievement than girls' achievement (in both 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. Feeling more confident in science was also associated with greater achievement gains in the subject among boys compared to girls.

#### Findings for grade 4

The results from the final two steps (steps 6 and 7) for the grade 4 mathematics and science models are presented in this section. Results for all steps are shown in Additional file 1: Tables S3–S7.

#### Mathematics

Table 10 provides the coefficients and model statistics for grade 4 mathematics from steps 6 and 7. Step 7, the final model, explains a substantial proportion of the observed variance in grade 4 mathematics achievement: 26% at level 1 (student level) and 71% at level 2 (class/school level), or 40% of the total observed variance.

As shown in step 6, the coefficient for 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, or due to the gender-segregated school system in Saudi Arabia, the achievement gap between schools attended by boys and schools attended by girls, 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<sup>10</sup> are significantly associated with student achievement, after holding other variables constant. 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% of an *SD*. Also, students with stronger early literacy and numeracy skills ( $B = 7.6$ ,  $p < 0.001$ ) and those who reported being bullied less frequently achieved higher mathematics scores ( $B = 4.9$ ,  $p < 0.001$ ) relative to other students. 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 ( $B = 54.9$ ,  $p < 0.05$ ), while with each additional year of teacher experience, students score 1.8 points higher in mathematics ( $p < 0.01$ ).

<sup>10</sup> The coefficient for 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 existing literature (e.g., Jha & Pouzevara, 2016).



**Table 10** 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)
<b>Student-level variables (reference category)</b>		<b>B (SE)</b>	<b>B (SE)</b>	<b>B (SE)</b>	<b>B (SE)</b>
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 <sup>a</sup>		4.91 (1.28)***	5.20 (1.26)***	5.78 (1.44)***	5.91 (1.44)***
<b>Class/school–level variables (reference category)</b>					
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)
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 min or less) <sup>b</sup>		4.63 (8.85)	9.68 (8.52)	– 7.30 (11.31)	– 5.47 (9.81)

**Table 10** (continued)

		Mathematics		Science	
		Step 6	Step 7	Step 6	Step 7
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 h)					
16–35 h		– 29.46 (11.69)*	– 22.62 (10.50)*	3.50 (14.09)	5.25 (11.73)
6–15 h		– 5.35 (13.24)	4.99 (12.26)	– 5.01 (12.01)	– 4.52 (10.92)
Less than 6 h		– 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) <sup>c</sup>		– 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)
<b>Interaction terms (reference category)</b>					
Gender*Safe and orderly schools			– 16.25 (5.13)**		– 15.61 (4.93)**
Gender*Student absenteeism (never or almost never)—once a week			22.88 (10.44)*		
Gender*Teacher age (40 years or older)—29 years or younger			89.46 (30.62)**		
Gender*literacy and numeracy readiness for school					– 6.29 (2.57)*
Gender*school location (urban)—suburban/medium size city or large town					75.76 (15.57)***
Fit statistics	Loglikelihood ( $H_0$ )	– 8549.92	– 8533.11	– 12381.76	– 12361.03
	AIC	17193.84	17166.23	24857.51	24822.05
	BIC	17442.83	17431.12	25123.27	25104.77

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

**Table 11** 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)
<b>Student-level variables (reference category)</b>		<b>B (SE)</b>	<b>B (SE)</b>	<b>B (SE)</b>	<b>B (SE)</b>
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 <sup>a</sup>		− 0.42 (0.91)	− 0.43 (0.91)	1.49 (1.21)	1.32 (1.21)
Time spent on mathematics/science homework (15 min or less) <sup>b</sup>		− 11.35 (3.98)**	− 11.40 (4.01)**	− 12.37 (3.99)**	− 11.94 (3.99)**
<b>Class/school-level variables (reference category)</b>					
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)
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 min or less) <sup>b</sup>		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)

**Table 11** (continued)

	Mathematics		Science	
	Step 6	Step 7	Step 6	Step 7
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) <sup>c</sup>	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 h)				
16–35 h	− 14.62 (8.62)	− 18.01 (8.67)*	− 3.53 (7.63)	− 3.66 (7.06)
6–15 h	− 0.91 (7.97)	− 2.98 (7.67)	− 0.71 (6.72)	− 4.61 (6.57)
Less than 6 h	− 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) <sup>d</sup>	− 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)
<b>Interaction terms (reference category)</b>				
Gender*School emphasis on academic success		− 6.43 (3.04)*		
Gender*Student confident in science				− 4.88 (1.91)*
Gender*Teacher age (40 years or older)—30–39 years				32.15 (10.20)**
Fit statistics	Loglikelihood ( $H_0$ )	− 12113.38	− 12109.54	− 13016.61
	AIC	24316.77	24311.07	26123.23
	BIC	24572.77	24572.76	26381.44

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

\*  $p < .05$

\*\*  $p < .01$

\*\*\*  $p < .001$

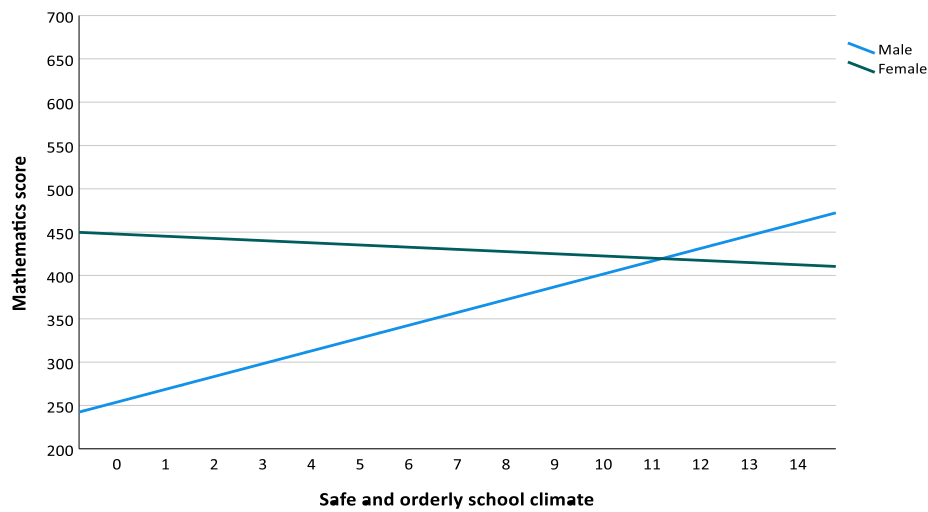
<sup>a</sup> Higher scores indicate less frequent bullying

<sup>b</sup> Other category: 16 min or more

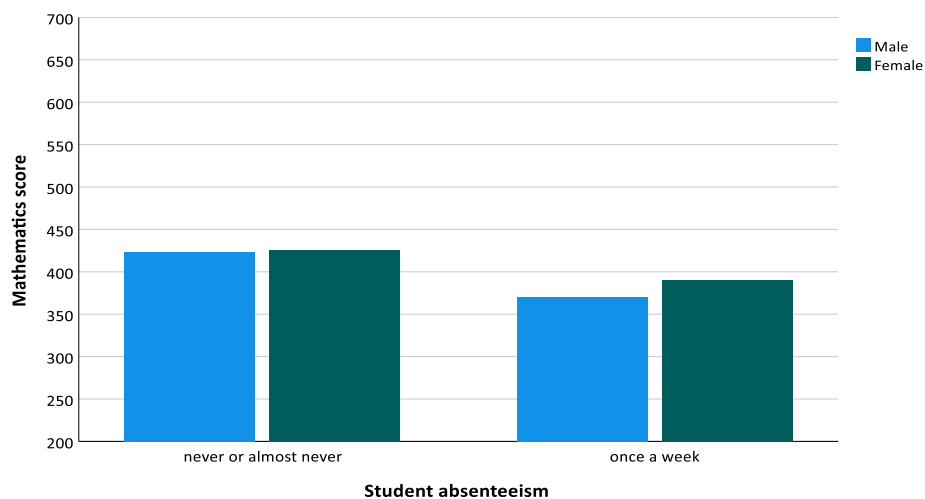
<sup>c</sup> Other category: ISCED levels 7 & 8 —master's or doctorate degree

<sup>d</sup> Other category: ISCED Level 6 — bachelor's or equivalent level

In step 7, the interactions between gender and each of the predictors were examined but only the results for interaction terms that are statistically significant are reported here. Overall, poor school climate tends to affect boys more negatively than girls. As shown in Fig. 2, the achievement gap 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.

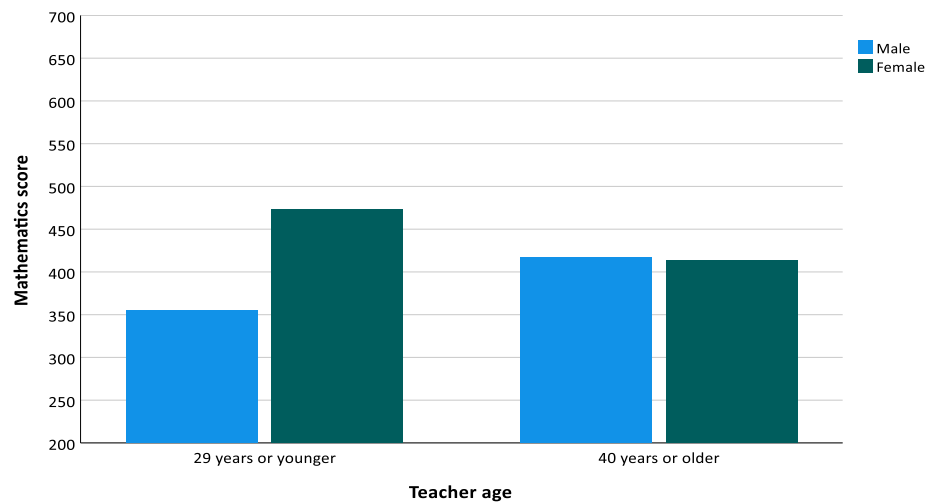


**Fig. 2** Interaction between gender and safe and orderly school climate on mathematics achievement, grade 4. The plot presents the predicted values from step 7 of the mathematics model for grade 4 (Table 10)



**Fig. 3** Interaction between gender and frequency of student absenteeism on mathematics achievement, grade 4. The plot presents the predicted values from step 7 of the mathematics model for grade 4 (Table 10). Only the reference category and the category for which a statistically significant interaction with gender was found are presented

As shown in Figs. 3 and 4, 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 (Fig. 3). 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 (Fig. 4).



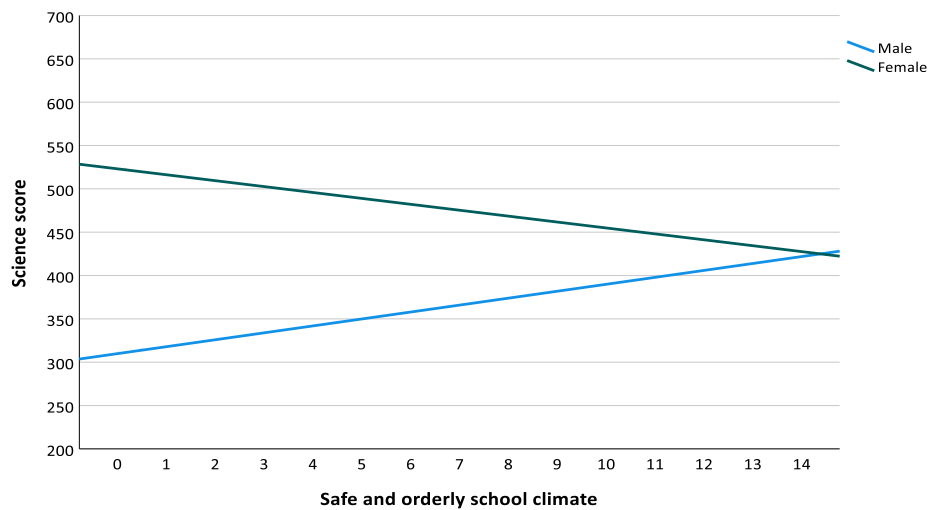
**Fig. 4** Interaction between gender and teacher age on mathematics achievement, grade 4. The plot presents the predicted values from step 7 of the mathematics model for grade 4 (Table 10). Only the reference category and the category for which a statistically significant interaction with gender was found are presented

### Science

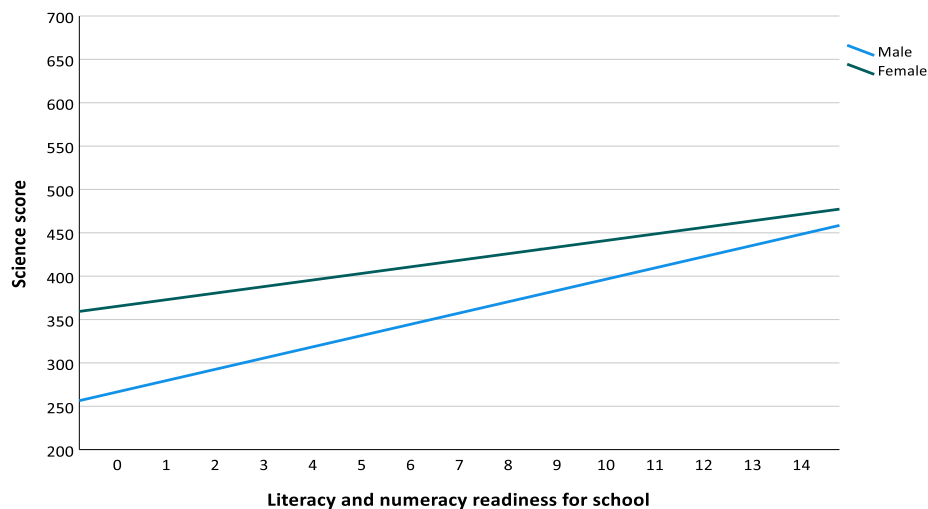
Table 10 also presents the coefficients and model statistics for grade 4 science. Similar to grade 4 mathematics, the results from steps 6 and 7 are presented. The final model explains a substantial proportion of the observed variance in grade 4 science achievement: 27% at level 1 and 65% at level 2, or 39% of the total observed variance. As shown in step 6, the coefficient for gender is positive and statistically significant ( $B=21.5$ ,  $p<0.05$ ). This coefficient is much smaller in magnitude compared to the basic model (step 1 in Table 9), which suggests that controlling for student, home, teacher, and school characteristics reduces the science achievement gap between grade 4 boys and girls, or due to the gender-segregated school system in Saudi Arabia, the achievement gap between schools attended by boys and schools attended by girls, by more than half. There is still, however, a significant unexplained gap between boys and girls in grade 4 science. 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 ( $B=7.6$ ,  $p<0.001$ ).

School location, school mean of home resources, teacher experience, age, and professional development are the class/school-level variables that were significantly associated with fourth-grade students' performance in science. After holding other variables constant, students in schools located in small towns or remote areas perform better in science than students in urban areas ( $B=60.8$ ,  $p<0.01$ ). 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



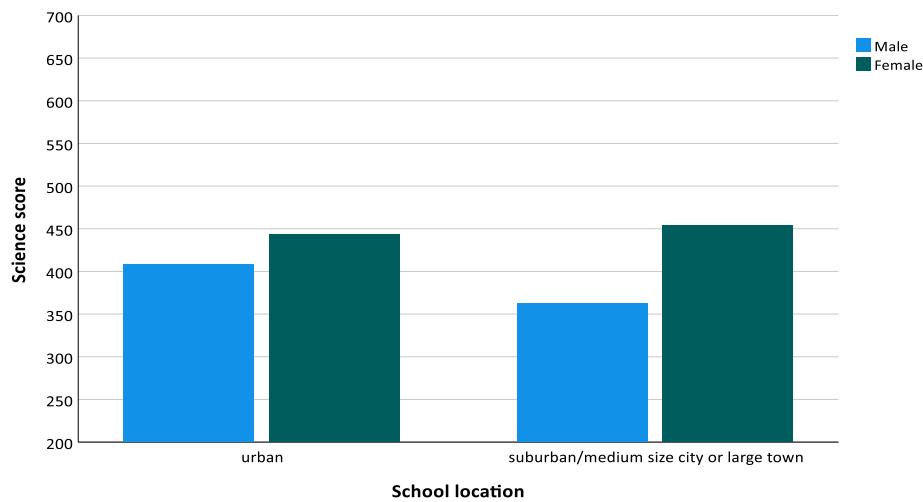


**Fig. 5** Interaction between gender and safe and orderly school climate on science achievement, grade 4. The plot presents the predicted values from step 7 of the science model for grade 4 (Table 10)



**Fig. 6** Interaction between gender and literacy and numeracy readiness for school on science achievement, grade 4. The plot presents the predicted values from step 7 of the science model for grade 4 (Table 10)

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 h in professional development in science ( $B = -35.3$ ,  $p < 0.05$ ). Consistent with grade 4 mathematics, results in science from step 7 show that poor school climate affects boys more negatively than girls (Fig. 5). 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 (Fig. 6), and girls in urban and suburban areas outperform boys (Fig. 7).



**Fig. 7** Interaction between gender and school location on science achievement, grade 4. The plot presents the predicted values from step 7 of the science model for grade 4 (Table 10). Only the reference category and the category for which a statistically significant interaction with gender was found are presented

### Findings for grade 8

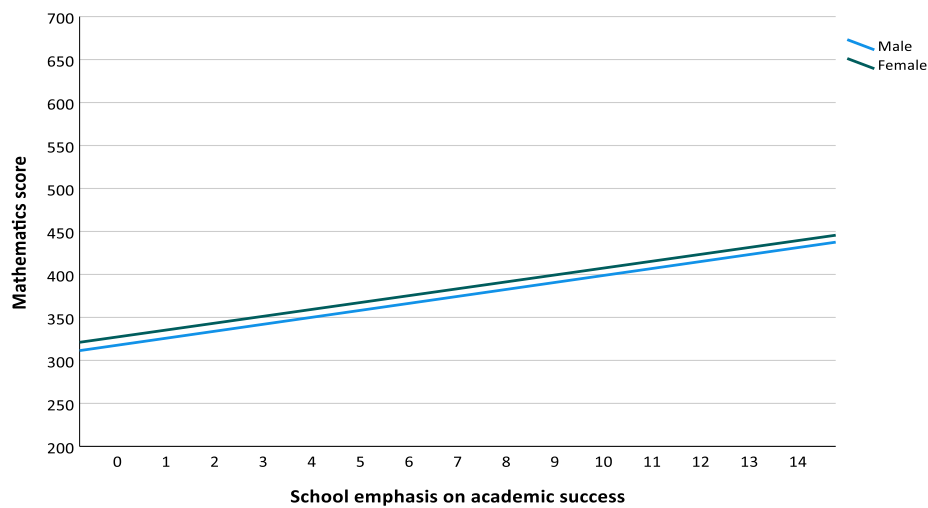
The results from the final two steps (steps 6 and 7) for the grade 8 mathematics and science models are presented in this section. Results for all steps are shown in Additional file 1 : Tables S3–S7.

#### Mathematics

Table 11 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 ( $B=16.3$ ,  $p<0.05$ ) than the gender difference recorded in step 1 ( $B=20.8$ , Table 9), 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% at level 1 and 74% at level 2, or 39% 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), feeling more confident in mathematics, lower liking of mathematics, and taking less time to complete mathematics homework (15 min or less). Class/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 statistically significant interaction with gender was observed for grade 8 mathematics. This interaction, involving schools' emphasis on academic success, is illustrated in Fig. 8. The interaction term indicates that boys' mathematics achievement increases to a greater degree than girls' achievement in schools with stronger emphasis on academic



**Fig. 8** Interaction between gender and school emphasis on academic success on mathematics achievement, grade 8. The plot presents the predicted values from step 7 of the mathematics model for grade 8 (Table 11)

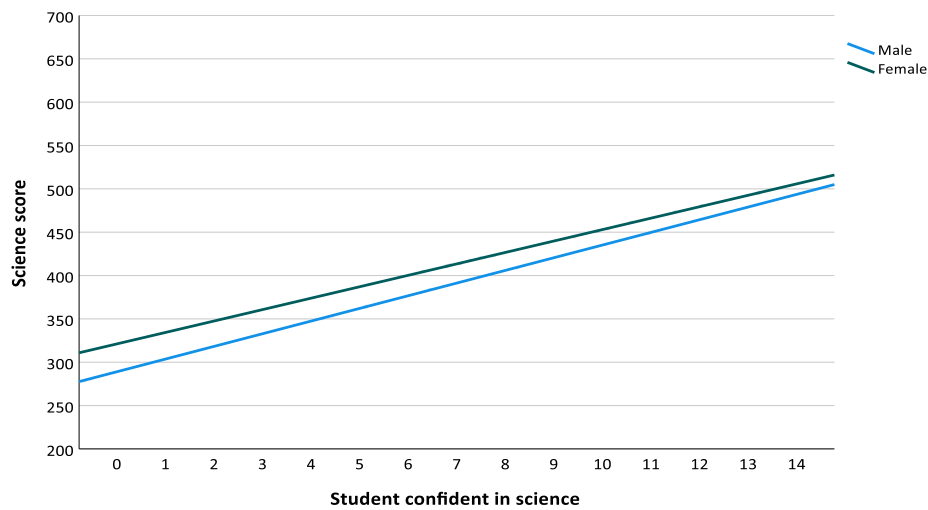
success, relative to schools with a weaker emphasis on academic success. However, though statistically significant, the magnitude of the interaction is small, as shown in Fig. 8. Given that this difference was not substantial, this finding should be interpreted with caution.

### Science

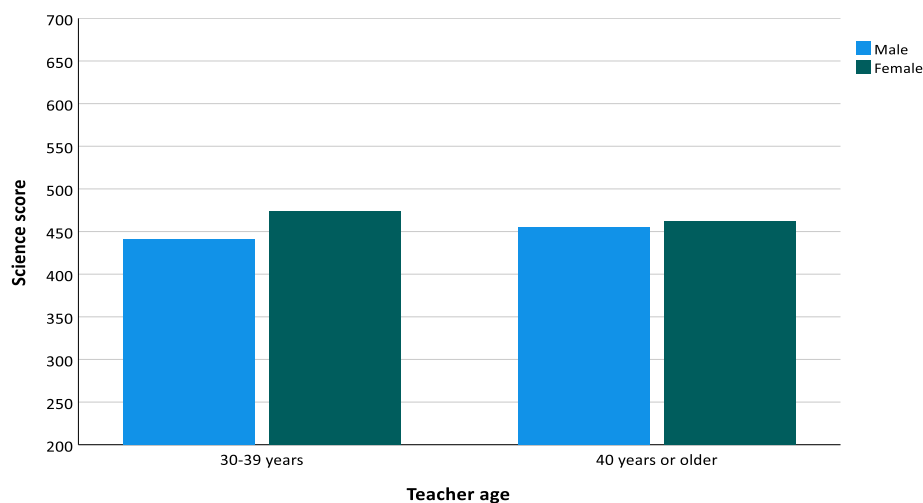
Table 11 also presents the coefficients and model statistics for grade 8 science. The gender difference in science achievement remains statistically significant and substantial ( $B = 39.9$ ,  $p < 0.001$ ) after the addition of the selected conceptual blocks of predictor variables. Nonetheless, the final model, including interactions, explains a substantial proportion of the observed variance in grade 8 science achievement: 27% at level 1 and 79% at level 2, or 41% 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), feeling more confident in science, and taking less time to complete science homework (15 min or less). Class/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 gender were observed. These interactions, involving the extent to which students feel confident in science and teachers' age, are illustrated in Figs. 9 and 10. The first interaction term indicates that reporting feeling more confident in science is 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



**Fig. 9** Interaction between gender and student confident in science on science achievement, grade 8. The plot presents the predicted values from step 7 of the science model for grade 8 (Table 11)



**Fig. 10** Interaction between gender and teacher age on science achievement, grade 8. The plot presents the predicted values from step 7 of the science model for grade 8 (Table 11). Only the reference category and the category for which a statistically significant interaction with gender was found are presented

higher in classes taught by younger teachers (ages 30–39 years old) than in classes taught by older teachers.

## Discussion

The results presented in this paper shed new light on the factors that are associated with mathematics and science achievement in Saudi Arabia, and on the factors that contribute to the large differences in achievement in these subjects between boys and girls, or due to the gender-segregated school system in Saudi Arabia, the differences in achievement between schools attended by boys and schools attended by 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 to identify key factors that can be considered as part of educators' and policy makers' efforts to raise achievement in elementary and intermediate schools in Saudi Arabia. Hence, summarized below are the most important findings of these analyses and some of the broader issues arising from them. In particular, attention is drawn next to the most robust findings with the clearest implications for educators and policy makers in Saudi Arabia.

### **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. 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 confident in mathematics or science, (d) being present at school more regularly, and (e) experiencing 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) feeling confident 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. As 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 four of the five being student-level factors.

### **Accounting for observed gender differences in achievement in Saudi Arabia**

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 other predictors. Although the gender difference in achievement is partially accounted for by the modeled variables—leading to a reduction in the 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 as test items on certain areas may have a high reading load (i.e., require a higher volume of reading or more complex reading skills). 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 have consistently shown that reading achievement is, on average, substantially lower in Saudi Arabia than in many other countries, both at grade 4 (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 an *SD* 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 tend to be more difficult for students to answer correctly than items with a lower reading load, and also that weaker readers are disproportionately disadvantaged by a higher reading load (Mullis et al., 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 proportion of girls).

The proposed importance of reading skills in underpinning mathematical and scientific achievement is consistent with the pattern of residual 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 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 class/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% of 15-year-old boys in Saudi Arabia reported repeating at least one grade, compared to 9.8% 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 and those who are struggling 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 may 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 et al., 2021; Corcoran et al., 2004). Differences in abilities between female and male teachers could be explained, in part, by the gender differences in returns to education across occupations (Cortes & Pan, 2018; World Bank, 2012). 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 more limited occupational opportunities for women outside the field of education (Carroll et al., 2021). Additionally, teaching is traditionally seen as a preferable profession for women in Saudi Arabia, which means that the competition among female graduates for teaching jobs 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 here have accounted for a high proportion of the observed variance in mathematics and science achievement in Saudi Arabia (ranging from 39 to 41% 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 for all students. However, the fact that the majority of variance in achievement (70–75%) 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.

### Limitations

The conclusions that can be drawn from the results reported in this paper are limited to being correlational in nature, as TIMSS and NALO are both cross-sectional studies. 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, 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 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 by the authors in the Saudi Arabia dataset for TIMSS 2019. TIMSS collects some data related to teachers and classroom practices but more detailed analysis on teacher quality would be possible with other studies explicitly 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. In these datasets, boys 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 in Saudi Arabia while controlling for teacher characteristics to a greater degree.

### Conclusions and implications

The findings of this study point to the relevance of the school climate in understanding 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 & Jeon, 2020) and, as indicated by the NALO data, greater support for learning for girls at home (Ridge & 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 their lesson material where possible while remaining alert to the effects of stereotypes (such as boys being more suited than girls to science and mathematics, or girls being more suited to reading) on 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 & 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 begin school with weaker early literacy and numeracy skills than girls. Moreover, 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 present in Saudi Arabia 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 models, two variables were found to be significantly associated with both mathematics and science achievement at both grade levels. These were students' feeling confident 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 than many others, should be considered carefully. The analyses presented here 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 (between the teacher or principal and the child's parents) and formal channels (formal communication between the school and the parents or, in more extreme cases, a state agency tasked 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 & 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,<sup>11</sup> boys in Saudi schools achieve much poorer outcomes than girls in both mathematics and science. Although holding higher qualifications does not necessarily imply a higher standard of teaching

---

<sup>11</sup> 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.

(Harris & 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.

### **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 dataset. 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 careers. In particular, considering the importance of literacy as a foundational skill (Gregory et al. 2021), the inclusion of an indicator of reading achievement would help to control for gender differences relating to literacy and would allow more fine-grained examination of mathematical and scientific proficiency. Among international assessments, data from PISA or from a joint TIMSS and PIRLS assessment (such as TIMSS/PIRLS 2011) 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, 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 examine factors associated with gender differences in single-gender compared to mixed-gender schools (see Fig. 1), particularly in, but not limited to, the MENA region.

## Abbreviations

CSEC	Caribbean Secondary Education Certification
ETEC	Education and Training Evaluation Commission
GDP	Gross Domestic Product
GER	Gross Enrollment Ratio
IEA	International Association for the Evaluation of Educational Achievement
IDB	International Database Analyzer
MENA	Middle East and North Africa
NALO	National Assessment of Learning Outcomes
OECD	Organisation for Economic Co-operation and Development
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
SPSS	Statistical Package for the Social Sciences
TALIS	Teaching and Learning International Survey
TIMSS	Trends in International Mathematics and Science Study

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40536-022-00141-9>.

**Additional file 1: Table S1.** Predictor variables from the TIMSS and NALO databases. **Table S2.** TIMSS cut-scores for categories of continuous indices, TIMSS 2019. **Table S3.** Equations for each step of the hierarchical two-level linear regression models. **Table S4.** Hierarchical two-level linear regression model for mathematics achievement, grade 4, TIMSS 2019. **Table S5.** Hierarchical two-level linear regression model for science achievement, grade 4, TIMSS 2019. **Table S6.** Hierarchical two-level regression model for mathematics achievement, grade 8, TIMSS 2019. **Table S7.** Hierarchical two-level regression model for science achievement, grade 8, TIMSS 2019.

## Acknowledgements

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

## Author contributions

All authors contributed to the design of the study. AC and VP analyzed and interpreted the data and together with ME, who managed the study, wrote all sections of the paper. NA and KA provided access to the national data, shared information about the national context as well as feedback in writing and in group discussions throughout the study, and contributed to the drafting of the paper. All authors read and approved the final manuscript.

## Funding

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

## Availability of data and materials

The TIMSS datasets analysed in this study are publicly available on the TIMSS website (<https://timssandpirs.bc.edu/timss-landing.html>). NALO data are not publicly available. Permission to use the NALO data was granted to the authors by the Education and Training Evaluation Commission (ETEC), Saudi Arabia. Requests to access NALO data should be made to ETEC.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

Received: 8 May 2022 Accepted: 14 November 2022

Published online: 16 December 2022

## References

- Angrist, N., Djankov, S., Goldberg, P. K., & Patrinos, H. A. (2021). Measuring human capital using global learning data. *Nature*, 592, 403–408. <https://doi.org/10.1038/s41586-021-03323-7>
- Autor, D., Figlio, D., Karbownik, K., Roth, J., & Wasserman, M. (2016). School quality and the gender gap in educational achievement. *American Economic Review*, 106(5), 289–295. <https://doi.org/10.1257/aer.p20161074>

- Autor, D., Figlio, D., Karbownik, K., Roth, J., & Wasserman, M. (2019). Family disadvantage and the gender gap in behavioral and educational outcomes. *American Economic Journal: Applied Economics*, 11(3), 338–381. <https://doi.org/10.1257/app.20170571>
- Bertrand, M., & Pan, J. (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>
- Booth, A. L., & Nolen, P. (2012). Gender differences in risk behaviour: Does nurture matter? *The Economic Journal*, 122(558), F56–F78. <https://www.jstor.org/stable/41418970>
- Brozo, W. G., Sulkunen, S., Shiel, G., Garbe, C., Pandian, A., & Valtin, R. (2014). Reading, gender, and engagement. *Journal of Adolescent & Adult Literacy*, 57(7), 584–593. <https://doi.org/10.1002/jaal.291>
- Buchmann, C., & DiPrete, T. A. (2006). The growing female advantage in college completion: The role of family background and academic achievement. *American Sociological Review*, 71(4), 515–541. <https://doi.org/10.1177/000312240607100401>
- Carroll, D., Parasnis, J., & Tani, M. (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, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Corcoran, S. P., Evans, W. N., & Schwab, R. M. (2004). Women, the labor market, and the declining relative quality of teachers. *Journal of Policy Analysis and Management*, 23(3), 449–470. <https://www.jstor.org/stable/3326261>
- Cornwell, C., Mustard, D. B., & Van Parys, J. (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources*, 48(1), 236–264. <https://doi.org/10.3368/jhr.48.1.236>
- Cortes, P., & Pan, J. (2018). Occupation and gender. In S. L. Averett, L. M. Argys, & S. D. Hoffman (Eds.), *The Oxford Handbook of Women and the Economy* (pp. 424–452). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190628963.013.12>
- DiPrete, T. A., & Buchmann, C. (2013). *The rise of women: The growing gender gap in education and what it means for American schools*. Russell Sage Foundation.
- DiPrete, T. A., & Jennings, J. L. (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, D. B., & Vogt Yuan, A. S. (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>
- Eisenkopf, G., Hessami, Z., Fischbacher, U., & Ursprung, H. W. (2015). Academic performance and single-sex schooling: Evidence from a natural experiment in Switzerland. *Journal of Economic Behavior & Organization*, 115, 123–143. <https://doi.org/10.1016/j.jebo.2014.08.004>
- Fortin, N. M., Oreopoulos, P., & Phipps, S. (2015). Leaving boys behind—Gender disparities in high academic achievement. *Journal of Human Resources*, 50(3), 549–579. <https://doi.org/10.3368/jhr.50.3.549>
- Fritz, C. O., Morris, P. E., & Richler, J. J. (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, L., Taha-Thomure, H., Kazem, A., Boni, A., Elsayed, M. A. A., & Taibah, N. (2021). *Advancing Arabic Language teaching and learning: A path to reducing learning poverty in the Middle East and North Africa*. World Bank. <https://openknowledge.worldbank.org/handle/10986/35917>
- Harris, D. N., & Sass, T. R. (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, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- IBM Corporation. (2020). *IBM SPSS Statistics for Windows* (No. 27). IBM Corporation.
- IEA. (2021). *Help manual for the IEA IDB Analyzer (Version 4.0)*. <https://www.iea.nl>
- Jackson, C. K. (2012). Single-sex schools, student achievement, and course selection: Evidence from rule-based student assignments in Trinidad and Tobago. *Journal of Public Economics*, 96(1–2), 173–187. <https://doi.org/10.1016/j.jpubeco.2011.09.002>
- Jha, J., & Pouezevara, S. (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*. United States Agency for International Development (USAID).
- Knoster, K. C. (2016). *Strategies for addressing student and teacher absenteeism: A literature review*. U.S. Department of Education, North Central Comprehensive Center.
- LaRoche, S., Joncas, M., & Foy, P. (2020). Sample design in TIMSS 2019. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report* (pp. 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, J., & DiPrete, T. A. (2012). School context and the gender gap in educational achievement. *American Sociological Review*, 77(3), 463–485. <https://doi.org/10.1177/0003122412440802>
- Mullis, I. V. S., Martin, M. O., Foy, P., & Drucker, K. T. (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).
- Mullis, I. V. S., Martin, M. O., & Foy, P. (2013). The impact of reading ability on TIMSS mathematics and science achievement at the fourth grade: An analysis by item reading demands. In M. O. Martin & I. V. S. Mullis (Eds.), *TIMSS and PIRLS 2011: Relationships among reading, mathematics, and science achievement at the fourth grade —Implications for early learning* (pp. 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](https://timssandpirls.bc.edu/timsspirls2011/downloads/TP11_Chapter_2.pdf)
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 international results in reading*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, and International Association for the Evaluation of Educational Achievement (IEA).

- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College, and International Association for the Evaluation of Educational Achievement (IEA).
- Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.
- Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice. *Theory and Research in Education*, 7(2), 133–144. <https://doi.org/10.1177/1477878509104318>
- Nilsen, T., Blömeke, S., Hansen, K. Y., & Gustafsson, J.-E. (2016). *Are school characteristics related to equity? The answer may depend on a country's developmental level* (Policy Brief No. 10). International Association for the Evaluation of Educational Achievement (IEA).
- OECD. (2015). *The ABC of gender equality in education: Aptitude, behaviour, confidence*. PISA, OECD Publishing. <https://doi.org/10.1787/9789264229945-en>
- OECD. (2019). *PISA 2018 results (volume I): What students know and can do*. PISA, OECD Publishing. <https://doi.org/10.1787/5f07c754-en>
- OECD. (2020). *Education in Saudi Arabia (Reviews of National Policies for Education)*. OECD Publishing. <https://doi.org/10.1787/76df15a2-en>
- OECD. (2021). *Positive, high-achieving students? What schools and teachers can do*. TALIS, OECD Publishing. <https://doi.org/10.1787/3b9551db-en>
- Page, E., & Jha, J. (Eds.). (2009). *Exploring the bias: Gender and stereotyping in secondary schools*. Commonwealth Secretariat. <https://doi.org/10.14217/9781848590427-en>
- Pahlke, E., Hyde, J. S., & Allison, C. M. (2014). The effects of single-sex compared with coeducational schooling on students' performance and attitudes: A meta-analysis. *Psychological Bulletin*, 140(4), 1042–1072. <https://doi.org/10.1037/a0035740>
- Pahlke, E., Hyde, J. S., & Mertz, J. E. (2013). The effects of single-sex compared with coeducational schooling on mathematics and science achievement: Data from Korea. *Journal of Educational Psychology*, 105(2), 444–452. <https://doi.org/10.1037/a0031857>
- Patrinos, H. A., & Angrist, N. (2019). *Harmonized learning outcomes: Transforming learning assessment data into national education policy reforms*. World Bank Blogs. <https://blogs.worldbank.org/opendata/harmonized-learning-outcomes-transforming-learning-assessment-data-national-education>
- Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., & Stringfield, S. (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, N., & Jeon, S. (2020). Father involvement and education in the Middle East: Geography, gender, and generations. *Comparative Education Review*, 64(4), 725–748. <https://doi.org/10.1086/710768>
- Rogers, H. F., & Vegas, E. (2009). *No more cutting class? Reducing teacher absence and providing incentives for performance*. World Bank Policy Research Working Paper No. 4847. <https://doi.org/10.1596/1813-9450-4847>
- Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher*, 39(2), 142–151. <https://doi.org/10.3102/0013189X10363170>
- Stromquist, N. (2007). *The gender socialization process in schools: A cross-national comparison (2008/ED/EFA/MRT/PI/71)*. Background paper for the Education for All Global Monitoring Report 2008, Education for All by 2015: Will we make it?
- Teixeira, P. J., Marques, M. M., Silva, M. N., Brunet, J., Duda, J. L., Haerens, L., La Guardia, J., Lindwall, M., Lonsdale, C., Markland, D., Michie, S., Moller, A. C., Ntoumanis, N., Patrick, H., Reeve, J., Ryan, R. M., Sebire, S. J., Standage, M., Vansteenkiste, M., Weinstein, N., Weman-Josefsson, K., Williams, G. C., & Hagger, M. S. (2020). A classification of motivation and behavior change techniques used in self-determination theory-based interventions in health contexts. *Motivation Science*, 6(4), 438–455. <https://doi.org/10.1037/mot0000172>
- von Davier, M., Gonzalez, E., & Mislevy, R. J. (2009). What are plausible values and why are they useful? *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, 2, 9–36.
- World Bank. (2012). *World development report 2012: Gender equality and development*. World Bank. <https://openknowledge.worldbank.org/handle/10986/4391>
- Younger, M., & Cobbett, M. (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>

## Publisher's Note

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