Gender Differences in Reading Achievement in Oman: Evidence from PIRLS 2021

Muhammad bin Saleh Al-Shizawi[1], Adil bin Hamdan bin Zayed Al-Rudaini[2], Fatma Rashid Salim Al Alyani[3], Sheikha bint Suhail Al-Sawwafi[1], Ali Salim Saif Almanei[2]

[1]Sultan Qaboos University, Sultanate of Oman, [2]Sohar University, Sultanate of Oman, [3]Ministry of Education, Sultanate of Oman

Abstract Girls outperform boys in reading in most education systems, especially in the Gulf. This study examined how much of Oman's Grade 4 gender gap in reading is statistically accounted for by home and school background, parent-reported literacy skills at school entry, and students' reading attitudes and school experiences, using PIRLS 2021 data from 5,321 students in 215 schools. Oaxaca-Blinder decompositions of three nested models, based on plausible values and jackknife standard errors, were supplemented by within-classroom, quantile, two-level and missing-data analyses. Girls scored 35.5 points higher than boys, a gap widest among lower achievers. For the 4,423 students with complete data, the three models accounted for 11.8%, 18.7%, and 49.7% of the 31.6-point gap. In the full sample, the most inclusive model accounted for 48.0% to 51.4%. Reading confidence, girls' lower exposure to bullying, and school-entry skills contributed most. These non-causal associations point to early literacy, reading confidence, and bullying as priorities for research and evaluation.

Keywords: *Gender gap, Gender inequality, Reading achievement, Gulf education, Oaxaca-blinder decomposition*

***Corresponding Author:**
Muhammad bin Saleh Al-Shizawi
m.alshizawi@squ.edu.om

Received: July 4, 2026
Revised: October 1, 2026
Accepted: October19, 2026
First Published: November 3, 2026

© 2026 Al-Shizawi1, Al-Rudaini, Al Alyani, Al-Sawwafi, and Almanei.
This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).

1. Introduction

Girls outperform boys in reading in most education systems that participate in international assessments (Mullis et al., 2023; Steinmann et al., 2023). In the Progress in International Reading Literacy Study (PIRLS) 2021, Grade 4 girls outscored boys by 18 points on average across 57 countries, and the difference was significant in all but five of them (Mullis et al., 2023). The advantage has persisted for decades, although country trends often depart from the international pattern (Steinmann et al., 2023).

The Gulf Cooperation Council (GCC) countries show some of the largest girls' advantages, although Qatar and the United Arab Emirates show smaller gaps than Bahrain, Kuwait, Oman, and Saudi Arabia (Al-Ubaydli &

https://doi.org/10.22034/ijscl.2026.2090444.4758

Alsyoof, 2025). Proposed explanations include the different climates of boys' and girls' schools (Basharat & Steinmann, 2025), but this regional evidence is largely descriptive. Decompositions that divide a reading gap into a part statistically accounted for by measured characteristics and a residual part have been reported for South Africa (Hofmeyr, 2022) and Morocco (Ibourk & El Aynaoui, 2025). Comparable published decompositions for the Gulf appear limited.

Published work on Oman addresses related but different questions. Ministry of Education records show girls ahead of boys in most subjects (Osman et al., 2015), and the girls' advantage in Grade 9 science was wider when boys reported lower self-efficacy and less positive attitudes (Al-Balushi et al., 2022). Other studies examined English reading and self-efficacy (Al Khamisi et al., 2016) and the alignment of Grade 4 Arabic textbooks with PIRLS (Alajmi & Al Balushi, 2024). Searches in 2026 combining "Oman", "PIRLS", "gender", "reading" and "decomposition" did not identify a published study that decomposes Oman's PIRLS reading gap or compares girls and boys within the same Omani classrooms. The gap is also a recognised policy concern. In Oman's first cycle of basic education (Grades 1 to 4), girls and boys are taught together by female teachers, whereas from Grade 5 they attend separate schools. The Ministry of Education also describes teacher training intended to narrow the difference between girls' and boys' reading achievement (Al Maskari et al., 2022). The coeducational first cycle is important for the analysis because girls and boys share classrooms. Therefore, the gap can be decomposed among students taught by the same teacher, which is not possible in systems with extensive single-sex schooling (Steinmann & Strietholt, 2026). Accordingly, this study asks three questions.

1. How large is the girls' advantage in Grade 4 reading in Oman, and how does it vary across the achievement distribution?
2. How much of the gap is statistically accounted for by home and school background, retrospectively reported literacy skills at school entry, and students' attitudes and school experiences, entered in that order?
3. Which factors make the largest contributions, and do the results hold within classrooms, across the distribution, and under alternative analytical choices?

2. Theoretical Framework

2.1. The Home Literacy Model

The home literacy model distinguishes informal literacy experiences, such as shared book reading, from formal experiences in which parents teach letters and words; in a five-year longitudinal study, parent teaching predicted early literacy skills, which in turn predicted reading in Grade 3 (Senechal & LeFevre, 2002). The model implies a sequence from home activities to the skills children bring to school and then to later reading. From this perspective, girls and boys may begin formal instruction with different literacy skills if they differ in the home experiences that precede it. The studies reviewed below are therefore organised along this sequence: first, whether home literacy experiences differ for girls and boys, and second, whether these experiences and the skills that follow from them predict later reading achievement.

Evidence on the first step indicates that home experiences do differ by sex. Parents engage more often in literacy activities with daughters than with sons, and these unequal home environments predicted unequal preliteracy outcomes (Hojen et al., 2022). In this Danish study of 8,469 preschool-age children, the home literacy environment explained part of the gender gap in language and preliteracy skills, although a direct association with gender remained. These early differences may have consequences well beyond school entry, since literacy enrichment in toddlerhood was related to later school engagement differently for girls and boys in a Quebec cohort of 1,997 children followed into adolescence (Kazandjian et al., 2025).

Evidence on the second step comes mainly from PIRLS. In PIRLS 2016 data from 1,025 Grade 4 learners in South Africa, reading books with the child before school and parents' participation in reading outside school hours were among the strongest predictors of reading literacy (Graham & Mtsweni, 2025). In Flanders, an active home reading climate was related to PIRLS 2021

IN PRESS

comprehension partly through students' reading enjoyment, motivation, and reading frequency (Claes et al., 2024). This indirect path suggests that the home environment also acts through motivational beliefs. In Morocco, pre-primary attendance and maternal education were among the observed factors associated with the girls' reading advantage (Ibourk & El Aynaoui, 2025). In Türkiye, however, parent-reported early literacy tasks were not associated with PIRLS 2021 achievement (Coshkun & Goktash, 2026), which suggests that the role of school-entry skills may differ between systems.

One possible source of such differences is the language in which children learn to read, which is particularly relevant in Arabic. Learning to read Standard Arabic after acquiring a spoken variety adds a linguistic demand that early home exposure may ease (Saiegh-Haddad, 2003), and among 110 Saudi children in Grades 2 to 4, spoken Arabic skills and decoding each contributed uniquely to the comprehension of Standard Arabic (Rakhlin et al., 2025). Taken together, these studies suggest that home literacy experiences and the skills children bring to school are plausible sources of the girls' advantage, but that their contribution varies between systems.

2.2. Expectancy-Value Theory

Expectancy-value theory holds that performance depends on beliefs about one's competence and on the value attached to a task (Wigfield & Eccles, 2000); confidence in reading reflects competence beliefs, and liking reading reflects task value. In this theory, such beliefs and values are not fixed traits but develop through children's experiences and through the beliefs and behaviour of parents, teachers and peers, including their gender-role stereotypes (Wigfield & Eccles, 2000). The reciprocal effects model adds that self-concept and achievement influence each other over time (Marsh & Craven, 2006), so confidence is both a possible source of the gap and partly a reflection of it. The studies reviewed below follow this line of reasoning, moving from the role of motivation in reading to the sources of sex differences in motivation, both in gendered beliefs and in students' school experiences.

Studies of reading motivation support the first part of the theory. Reading motivation predicted comprehension beyond decoding and vocabulary, and its predictive role increased from Grade 2 to Grade 4 among 420 Chinese children (Li et al., 2024). In PIRLS 2021 data, reading self-efficacy or confidence was among the strongest student-level correlates of achievement (Chalishkan & Goktenturk, 2026; Thomure et al., 2025; Yalchin, 2026). In a structural model of 6,032 Turkish students, for example, reading self-efficacy had the strongest effect among personal, family, and school variables (Chalishkan & Goktenturk, 2026). Boys' motivation may also depend more on the texts they read: among 514 Grade 4 students, boys' interest depended on the topic, protagonist, and difficulty of a text, whereas girls' interest depended only on difficulty, and interest was more strongly related to comprehension for boys (Lepper et al., 2022).

The theory explains sex differences in these beliefs through socialisation, in which reading is often framed as a feminine activity. Gender identity accounted for differences in reading motivation better than sex itself among 182 children aged 8 to 11 (McGeown et al., 2012); teachers' stereotypes were linked to boys' lower reading self-concept from Grade 5 to Grade 6 in a study of 54 teachers and 1,358 students (Retelsdorf et al., 2015); and agreement that reading is for girls was associated with boys valuing reading less (Espinoza & Strasser, 2020; Espinoza Catalan et al., 2025). In an experiment with 5,224 Grade 5 students in China, a message that girls read better lowered boys' scores without significantly affecting girls' scores (Huang et al., 2025).

Such expectations may also be conveyed through students' daily experiences with teachers and peers. Teachers have been linked to the size of the reading gap (Aucejo et al., 2022). In that study, differences between teachers in how much they raised boys' achievement relative to girls' could account for a large share of the gap, and these differences were related to gaps in how girls and boys perceived the same teacher's practices. Experiences with peers may affect both the value students attach to school and their achievement. Bullying victimisation is associated with lower achievement: across 79 countries in PISA 2018, bullied students scored lower in reading, mathematics and science, and in mathematics the association was stronger for boys (Gimenez et al., 2024). In Türkiye, bullying was among the student-level predictors of PIRLS 2021 achievement (Coshkun & Goktash, 2026), and in the United Arab Emirates, bullying predicted Grade 4 students' attitudes towards school in TIMSS

2023, with cyberbullying also predicting their mathematics and science achievement (Opoku, 2026). A sense of belonging at school shows a similar link with achievement. School belonging and achievement have been found to affect each other, and in a longitudinal US sample followed through Grades 4 and 5, belonging predicted later achievement more strongly than the reverse (Sakellariou, 2025). Engagement in reading lessons and disorderly behaviour describe the classroom climate students experience. In Oman's mixed classrooms, girls and boys share school-level conditions, but individual students may still experience the same classroom differently.

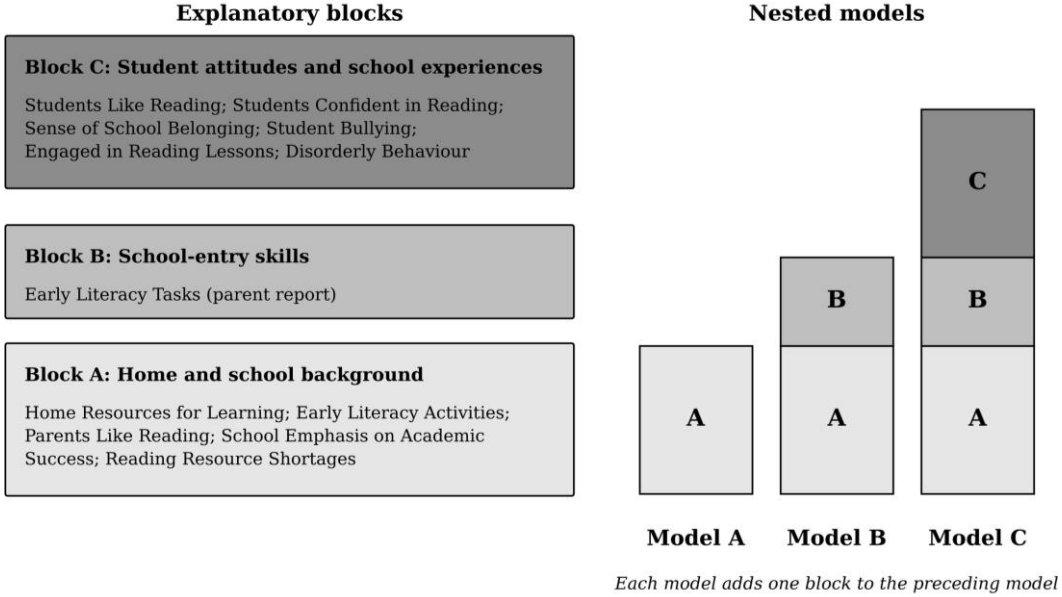
2.3. Previous Decompositions, Research Gap, and the Present Study

Few studies have examined these strands together by decomposing the girls' reading advantage. In South Africa, where the girls' advantage is among the largest observed, Oaxaca-Blinder decompositions of PIRLS data linked part of the gap to girls' faster progress through the early grades and to their higher reading self-efficacy, enjoyment, and engagement in reading lessons (Hofmeyr, 2022). In Morocco, observed characteristics accounted for about a fifth of the reading gap in PIRLS 2016 and 2021 (Ibourk & El Aynaoui, 2025). Across 56 education systems, changes in the gender gap in reading for enjoyment were related to changes in the gender gap in reading achievement (Steinmann & Strietholt, 2026). In the Gulf, the reading gap is part of a reverse gender divide in education (Ridge, 2014). Girls outperform boys across grades and subjects in TIMSS and PIRLS, although the gap is smaller in Qatar and the United Arab Emirates and narrowed after the COVID-19 pandemic (Al-Ubaydli & Alsyoof, 2025). In the United Arab Emirates, gender, reading confidence, and reading engagement were significant student-level predictors of PIRLS 2021 achievement (Thomure et al., 2025), and in Oman, students' self-efficacy and attitudes were related to the girls' advantage on higher-level science items (Al-Balushi et al., 2022). These studies either describe the Gulf gap or examine predictors of achievement. No published study was identified that decomposes Oman's PIRLS reading gap or compares girls and boys taught in the same Omani classrooms.

The present study combines the two strands reviewed above to address this gap. A variable can account for part of the gap only if girls and boys differ on it and it is associated with achievement. Home resources and school characteristics are expected to be associated with achievement but to differ little by sex; together with early literacy activities, they form the first block (Model A). The home literacy model places school-entry skills between home activities and later reading, so they form the second block (Model B). According to the socialisation perspective, attitudes, confidence and school experiences should differ by sex; according to the reciprocal effects model, they may also be partly consequences of achievement. They therefore enter last (Model C). Figure 1 shows the three nested models.

Figure 1

Nested Explanatory Blocks in Models A, B and C



3. Methodology

3.1. Participants

The data come from the public-use PIRLS 2021 International Database (Fishbein et al., 2024). PIRLS uses a two-stage stratified cluster design: schools are sampled with probability proportional to size within strata, and one or more intact Grade 4 classes are selected in each school, with all their students assessed. The total student weight (TOTWGT), the product of the school, class, and student weighting factors and nonresponse adjustments, allows estimates to represent the national population and is used in every estimate.

The Omani sample contains 5,321 students (2,687 girls and 2,634 boys; 50.4% girls after weighting) in 215 schools and 221 classes; the students' mean age was 9.8 years (SD = 0.3). Overall, 72.8% of students always or almost always spoke Arabic, the language of the test, at home; among those with parent data, 42.0% had a university-educated parent and 84.6% had attended pre-primary education. Of the 219 teachers who reported their sex, 218 were women. Of the 221 classes, 219 contained both girls and boys.

The main decomposition uses the 4,423 students (83.1%; 2,291 girls and 2,132 boys) with valid data on every variable of Model C, so that all nested models refer to the same students. These students retain their original weights. The 898 excluded students were more often boys (56.5% vs. 48.3%) and scored 62.1 points lower (SE = 6.4), although their girls–boys gap did not differ significantly from that of the analytic sample. The complete-data estimates therefore describe the retained students, whose gap (31.6 points) is smaller than that of the full sample (35.5 points). Four full-sample analyses test whether the exclusions change the conclusions (Section 3.3). The comparison systems are Bahrain (5,208 students), Qatar (5,258), Saudi Arabia (4,778), and the United Arab Emirates (27,448). Because PIRLS 2021 was administered during the COVID-19 pandemic (Kennedy & Strietholt, 2023), the analysis is confined to one cycle.

3.2. Instruments

Reading achievement is measured by the PIRLS 2021 assessment of literary and informational reading and reported as five plausible values on a scale with a centrepoint of 500 and a standard deviation of 100 (Mullis et al., 2023). The gap is the girls' mean minus the boys' mean. Students' sex is recorded as girl or boy; the positive difference in reading is called the gender gap or the girls' advantage, and "sex differences" refers to differences on the explanatory scales.

The explanatory variables are twelve scales from the student, home (parent), and school (principal) questionnaires (Table 1). PIRLS constructed each scale with the Rasch partial credit model after reverse-coding negatively worded items, scored respondents who answered at least two items, and set an international centrepoint of 10 and a standard deviation of 2 (Yin & Reynolds, 2023). The scores were used without modification, and no items were removed; omitted or invalid codes were set to missing. The Early Literacy Tasks scale is based on parents' reports made when the child is in Grade 4 and refers to skills at the start of Grade 1, so "school-entry skills" is shorthand for a retrospective parent report.

Table 1
Explanatory Variables by Block

Block	Scale (variable)	Respondent; items	Higher values indicate	α
A. Home	Home Resources for Learning (ASBGHRL)	Student and parent; 5	More learning resources	.45
A. Home	Early Literacy Activities Before School (ASBHELA)	Parent; 9	More frequent early literacy activities	.71
A. Home	Parents Like Reading (ASBHPLR)	Parent; 9	More positive parental attitudes to reading	.78
A. School	School Emphasis on Academic Success (ACBGEAS)	Principal; 12	Stronger academic emphasis	.90

Block	Scale (variable)	Respondent; items	Higher values indicate	α
A. School	Instruction Affected by Reading Resource Shortages (ACBGRRS)	Principal; 13	Instruction less affected by shortages	.96
B. School	Early Literacy Tasks (ASBHELT)	Parent; 6	Better literacy skills at the start of Grade 1, as reported by parents	.90
C. Student	Students Like Reading (ASBGSLR)	Student; 10	More positive attitudes to reading	.74 / .78
C. Student	Students Confident in Reading (ASBGSCR)	Student; 6	Higher confidence in reading	.66 / .69
C. Student	Sense of School Belonging (ASBGSSB)	Student; 5	Stronger sense of belonging	.71 / .79
C. Student	Student Bullying (ASBGSB)	Student; 10	Less frequent bullying	.90 / .92
C. Student	Engaged in Reading Lessons (ASBGERL)	Student; 9	Higher engagement in reading lessons	.83 / .85
C. Student	Disorderly Behaviour in Reading Lessons (ASBGDRL)	Student; 5	Less disorderly behaviour	.80 / .80

Note. PIRLS Rasch scale scores (centrepoint 10, standard deviation 2). Items use three- to five-point frequency or agreement formats. Model A contains block A, Model B adds block B, and Model C adds block C. α = Cronbach's alpha in Oman (girls/boys for student scales; schools for principal scales); the low alpha of Home Resources for Learning is discussed below the table.

Reliability was acceptable for most scales; it was lowest for Home Resources for Learning (.45) and, among the student scales, for confidence (.66 for girls and .69 for boys; PIRLS reports .68 for Oman). The low alpha partly reflects the heterogeneous indicators of this index, but any measurement error attenuates its coefficient and decomposition contribution. Corrected for a reliability of .45, this contribution rose from 1.4 to 3.5 points (*SE* = 1.7, unadjusted p = .04). It remained small because girls and boys differed little in home resources (*SD* = 0.07); the uncorrected 1.4 points is a lower bound, and the non-significance of this contribution after multiple-testing adjustment (Section 4.3) does not show that home resources contribute nothing. As validity evidence, all twelve scales correlated positively with reading, with confidence correlating most strongly (r = .35).

Measurement invariance across sex was tested for the six student scales by multi-group confirmatory factor analysis with robust maximum likelihood and normalised weights. Metric (equal loadings) and scalar (equal loadings and intercepts) models were each compared with the preceding, less constrained model using robust fit indices, and invariance was accepted only if both ΔCFI $\geq -.01$ (Cheung & Rensvold, 2002) and ΔRMSEA $\leq .015$ (Chen, 2007). Both criteria were met at the metric and scalar levels for all six scales (Table 2; ΔCFI −.009 to .000, ΔRMSEA −.013 to .003), so sex differences in scale means can be interpreted as differences in the constructs rather than in item functioning. The largest decrease in CFI (−.009, liking reading, scalar step) was close to the cut-off. Because the criteria concern changes in fit, they assume an adequate configural model, and this assumption was weakest for confidence. The confidence model fitted only moderately, even with a method factor for its four negatively worded items (CFI = .922, RMSEA = .102), and a one-factor model fitted poorly (CFI = .847). These items, which include comparisons with classmates and with other subjects, therefore appear to share variance beyond general confidence (omega = .73 for girls and .72 for boys).

Table 2

Measurement Invariance across Sex of the Student Scales

Scale	CFI	RMSEA	ΔCFI	ΔRMSEA	ΔCFI	ΔRMSEA
Students Like Reading	.936	.064	−.001	−.003	−.009	.000
Students Confident in Reading	.922	.102	.000	−.013	−.001	−.007

Scale	CFI	RMSEA	ΔCFI	ΔRMSEA	ΔCFI	ΔRMSEA
Sense of School Belonging	.993	.036	.000	−.004	−.004	.003
Student Bullying	.957	.079	.000	−.004	−.001	−.003
Engaged in Reading Lessons	.973	.052	−.001	−.003	−.001	−.002
Disorderly Behaviour in Reading Lessons	.995	.038	.000	−.005	−.002	.000

Note. CFI and RMSEA: configural model; first ΔCFI and ΔRMSEA: metric against configural model; second: scalar against metric model. Negative ΔRMSEA indicates better fit of the more constrained model. n = 4,327 to 4,802 students per scale. The confidence model includes a method factor for its negatively worded items.

3.3. Procedure

The analysis follows the research questions: weighted means, Cohen's d, benchmark proportions and percentiles describe the gap (RQ1); nested decompositions measure how much of it the three blocks account for (RQ2); and detailed contributions with within-classroom, quantile, two-level and sensitivity analyses identify the leading factors and test their stability (RQ3).

The decomposition follows Blinder (1973) and Oaxaca (1973). For each group g (girls, G; boys, B), reading is modelled by weighted least squares as $Y_{ig} = X'_{ig}\beta_g + \varepsilon_{ig}$, and the raw gap $\Delta = \overline{Y}_G - \overline{Y}_B$ is divided using a pooled reference vector β^P:

(1)

$$\Delta = E + U, \qquad E = \left(\overline{X}_G - \overline{X}_B\right)'\beta^P, \qquad U = \overline{X}'_G\left(\hat{\beta}_G - \beta^P\right) + \overline{X}'_B\left(\beta^P - \hat{\beta}_B\right).$$

(2)

$$E_k = \left(\overline{X}_{Gk} - \overline{X}_{Bk}\right)\beta_k^P, \qquad E = \sum_k E_k.$$

The explained part E is described as the share "statistically accounted for," and each E_k as a "decomposition contribution"; neither implies a causal effect. The unexplained part U is a residual and cannot be interpreted as discrimination, unequal treatment, or an effect of teacher behaviour. The pooled reference is the primary specification because using the girls' or the boys' coefficients would treat one group's associations as the norm for both (Neumark, 1988). Including the girl indicator in the pooled regression keeps intercept differences out of E (Fortin, 2008; Jann, 2008). With this reference, E evaluates sex differences in characteristics using associations common to both groups, and U equals the girl coefficient of the pooled regression; the girls' and boys' references are reported as sensitivity analyses. Model A is treated as the most conservative and Model C as the most inclusive statistical account, because several student measures may be consequences as well as correlates of achievement. The Model C contributions summed by block are independent of the order in which blocks are added (Gelbach, 2016), and changes in E between models are tested with the variance estimator described below.

The sampling design is incorporated into every estimate in four steps (Foy & Almaskut, 2023). (a) For each plausible value, the group means, the girls', boys', and pooled coefficients, and all components are computed with the total weight. (b) The computation is repeated with each of 220 jackknife replicate weights, formed by doubling and zeroing the weights of the paired schools in each of the 110 zones, and half the sum of squared deviations gives the sampling variance. (c) The five estimates are averaged. (d) The between-plausible-value variance, multiplied by $(1 + 1/M)$ with M = 5, is added to the average sampling variance. This procedure reproduces the published Oman results exactly: 429 (*SE* = 3.7) overall, 447 (*SE* = 4.2) for girls and 412 (*SE* = 4.1) for boys (Mullis et al., 2023).

The within-classroom decomposition subtracts the weighted class means from achievement and each explanatory variable. The resulting gap is a weighted sum of within-classroom girls–boys differences, identified from the 218 classes that still contained both girls and boys after students with incomplete

IN PRESS

data were excluded (219 mixed-sex classes in the full sample); school variables, constant within classes, drop out. Class means are computed with the total weight, and the demeaned data then pass through steps (a) to (d) with the same replicate weights and plausible values, so standard errors are comparable with those of the main analysis. The specification removes everything students in a class share, such as the teacher's general effectiveness, but not differential responses of teachers or classmates to girls and boys, unmeasured individual characteristics, or the role of teacher sex.

The quantile decomposition uses recentred influence functions (Firpo et al., 2009) at the 10th, 25th, 50th, 75th and 90th percentiles, which cover both tails, the quartiles and the median, and at which density estimates remain stable; weighted quantiles, densities and influence functions are recomputed for each plausible value and replicate weight, so standard errors again follow steps (a) to (d). Percentiles are group-specific, results are interpreted in score points, and a change in the gap across percentiles is distinguished from a change in an individual contribution.

Two-level models with students in schools, fitted to the same 4,423 complete-data students, show how far each block reduces the conditional girl coefficient. Because the jackknife replicate weights are constructed for single-level estimators, these models use the design weights of each level instead: the school weight at level 2 and the conditional weight of the student within the school at level 1, scaled to the school sample size (Rabe-Hesketh & Skrondal, 2006). The models are estimated by weighted pseudo-maximum likelihood, and the fixed effects, variance components, and their school-level cluster-robust standard errors are combined across plausible values by Rubin's rules. These standard errors reflect clustering and unequal weights but not the pairing of schools in jackknife zones, so they are not strictly comparable with the jackknife standard errors. The change in the girl coefficient is also a different estimand from the decomposition shares; the two-level results therefore serve only as a check on direction and approximate size. Because variance components depend on weight scaling, a sex-only random-slope model was estimated with scaled weights, unscaled weights, and no weights.

Missing data were handled by complete-case analysis and examined in four ways: multiple imputation by predictive mean matching (van Buuren & Groothuis-Oudshoorn, 2011) and inverse probability weighting, both assuming that missingness depends only on observed variables; missing-value indicators, a descriptive method that can bias coefficients (Jones, 1996); and a delta adjustment that shifts boys' imputed values by up to 0.5 standard deviations to represent departures from that assumption. Further sensitivity analyses varied the reading purpose, the school block and the background measures, omitted confidence, and corrected confidence and school-entry skills for measurement and recall error. Models A to C were estimated with missing-value indicators for the four other GCC systems, because parent data were missing for up to 49.8% of students in these systems.

P values for the 69 variable-level contributions from Models A to C, the within-classroom model, and the 10th, 50th, and 90th percentile models were adjusted with the Benjamini and Hochberg (1995) procedure; the 24 GCC contrasts, which compare independent national samples, form a separate family. "Adjusted *p*" denotes these values; all other *p* values are unadjusted. Analyses were run in R 4.5.2 (R Core Team, 2025) with base R code, mice 3.19.0, lavaan 0.6.21 (Rosseel, 2012) and WeMix 4.0.4 (Bailey et al., 2024).

4. Results

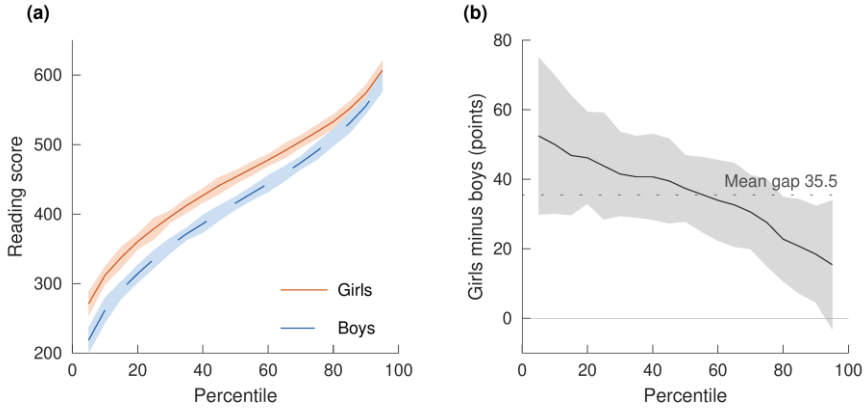
4.1. RQ1: Size and Distribution of the Gap

Girls scored 447.1 points (*SE* = 4.2) and boys 411.6 (*SE* = 4.1), a gap of 35.5 points (*SE* = 3.8, $p < .001$; $d = 0.33$). The gap was largest in the lower part of the distribution (Figure 2): 50.0 points (*SE* = 10.1) at the 10th percentile, 37.4 (*SE* = 4.8) at the median, and 18.5 (*SE* = 7.0) at the 90th percentile. Overall, 44.5% of boys and 31.2% of girls scored below the Low International Benchmark. Girls also scored higher in the four other GCC systems ($p < .001$ in each); Oman's gap was smaller than Bahrain's (adjusted $p = .044$), larger than Qatar's (adjusted $p = .006$), and similar to those of Saudi Arabia and the United Arab Emirates (Table 4). In answer to RQ1, the advantage was about a third of a standard deviation and largest among the weakest readers.

IN PRESS

Figure 2

Reading Scores of Girls and Boys at Each Percentile (a) and the Girls–boys Difference (b) (Oman, Full Sample), with 95% Confidence Intervals



4.2. RQ2: Gap Statistically Accounted for by the Three Blocks

Girls had significantly higher means than boys on ten of the twelve scales, particularly on the student measures; the differences in home resources were small, and the school measures did not differ. In the complete-data sample, home and school background accounted for 3.7 of the 31.6 points (11.8%; Table 3). Adding school-entry skills raised the explained part by 2.2 points (*SE* = 0.7, p = .002) to 18.7%, while the contribution of early literacy activities fell from 1.5 to 0.3 points. Adding the student block raised the explained part by a further 9.8 points (*SE* = 1.5, p < .001) to 15.7 points (49.7%), and, across the three models, the unexplained part fell from 27.9 to 15.9 points. In the two-level models of the complete-data sample, the girl coefficient (34.9 points) declined by 8.3%, 13.5% and 41.1% in Models A, B and C, relative to the sex-only model. In answer to RQ2, background accounted for about a tenth of the gap, school-entry skills raised this to about a fifth, and students' attitudes and experiences to about half.

Table 3

Oaxaca–Blinder Decomposition of the Reading Gap (Oman, Complete-Data Sample)

Component	Model A	Model B	Model C
Gap (girls minus boys)	31.64 (3.91)	31.64 (3.91)	31.64 (3.91)
Statistically accounted for (explained)	3.73 (1.16)	5.90 (1.32)	15.71 (1.85)
Not statistically accounted for (unexplained)	27.91 (3.63)	25.74 (3.57)	15.93 (3.22)
Percentage statistically accounted for	11.8 (3.4)	18.7 (3.9)	49.7 (5.6)
Home Resources for Learning	1.77 (0.86)	1.61 (0.79)	1.40 (0.69)
Early Literacy Activities	1.50 (0.45)[a]	0.30 (0.32)	0.25 (0.32)
Parents Like Reading	0.43 (0.27)	0.05 (0.21)	−0.12 (0.19)
School Emphasis on Academic Success	0.02 (0.24)	0.02 (0.27)	0.02 (0.25)
Reading Resource Shortages	0.01 (0.08)	0.00 (0.07)	−0.01 (0.07)
Early Literacy Tasks (school entry)		3.93 (0.87)[a]	3.09 (0.73)[a]
Students Like Reading			0.13 (0.58)
Students Confident in Reading			5.45 (0.90)[a]
Sense of School Belonging			1.14 (0.89)
Student Bullying			4.72 (0.90)[a]
Engaged in Reading Lessons			−0.15 (0.49)
Disorderly Behaviour			−0.21 (0.32)

Note. Score points with standard errors (jackknife and plausible-value variance) in parentheses; N = 4,423; pooled reference. Contributions are statistical associations, not causal effects. Scale names are shortened; see Table 1.
[a]Adjusted p < .05.

IN PRESS

4.3. RQ3: Largest Contributions and Their Stability

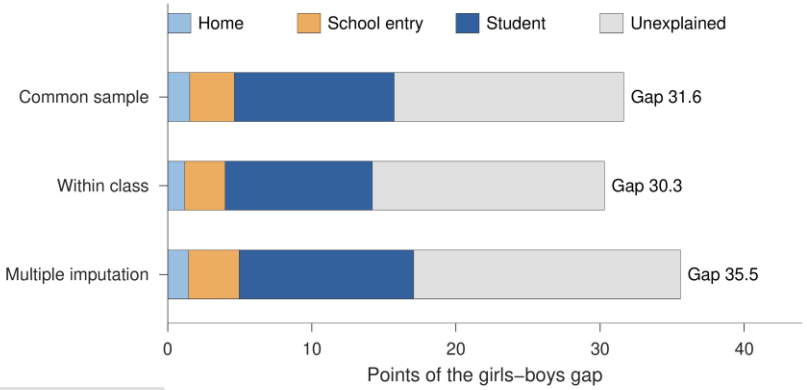
In the complete-data Model C, confidence (5.5 points) and bullying (4.7) made the largest contributions, followed by school-entry skills (3.1); the contribution of home resources (1.4) was not significant after adjustment (adjusted p = .14). By block, students' attitudes and experiences accounted for 35.0% of the gap, school-entry skills for 9.8%, home background for 4.8%, and the school block for 0.0%. After adjustment, confidence and bullying were significant in every model in which they appeared, and school-entry skills in every model except the 90th percentile model.

Within classrooms in the complete-data sample, the gap was 30.3 points, of which 47.0% (*SE* = 5.2) was accounted for, again mainly by confidence (5.2), bullying (3.8) and school-entry skills (2.8; Figure 3). Across the distribution (Figure 4), the gap fell from 43.3 points at the 10th percentile to 17.7 at the 90th, whereas the explained part fell only from 18.6 to 12.0 points. The narrowing therefore came mainly from the unexplained part, which was not significant at the 75th and 90th percentiles. Confidence contributed similarly at every percentile (4.5 to 6.2 points), whereas the contributions of school-entry skills (4.5 to 1.3) and bullying (6.2 to 3.1) declined towards the top.

In the sex-only random-slope model, the girls' advantage varied between schools with a standard deviation of 12.1 points (*SE* = 8.8; 95% CI 0 to 23.8) around a mean of 34.6. If school slopes were normally distributed, the advantage would lie between about 11 and 58 points in 95% of schools. Because the confidence interval includes zero and the estimate depended strongly on the weights (19 points unweighted, 39 with unscaled weights), the data do not establish how much the gap varies between schools.

Figure 3

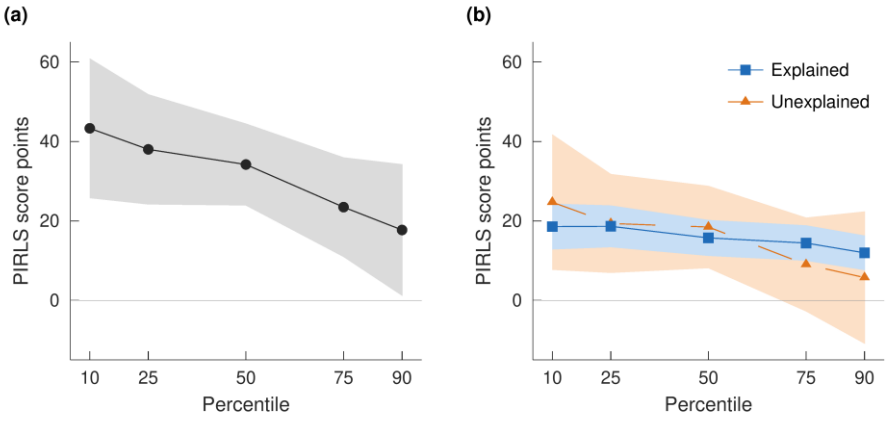
Block Contributions and the Unexplained Part of the Gap in Three Specifications (Oman, Model C)



Note. The school block (0.0 points) is not shown.

Figure 4

Girls-Boys Gap (a) and Its Explained and Unexplained Parts (b) at Five Percentiles (Oman, Complete-Data Sample, Model C), with 95% Confidence Intervals



In the complete-data sample, the explained share was 51.4% with the girls' and 48.4% with the boys' coefficients, compared with 49.7% with the pooled reference; confidence, bullying, and school-entry skills were significant under all three references, although the confidence contribution ranged from 4.6 to 6.3 points and belonging was significant only with the girls' coefficients. In the full sample of 5,321 students, the Model C explained share was 48.0% with multiple imputation, 51.4% with missing-value indicators (coincidentally equal to the girls'-reference share), 49.0% with inverse probability weighting, and 40.2% to 55.3% when boys' imputed values were shifted by ±0.5 standard deviations, with the same three leading factors throughout. Correcting confidence for unreliability raised its contribution to between 9.3 and 11.3 points, correcting school-entry skills for recall error raised their contribution to between 4.1 and 5.0 points, and omitting confidence lowered the complete-data explained share only to 43.9%.

In the GCC comparison, based on full samples with missing-value indicators, the explained share was lowest in Oman (51.4%; Table 4). Bullying contributed significantly less in Oman than in Bahrain, Saudi Arabia and the United Arab Emirates, and school-entry skills contributed more than in Qatar, Saudi Arabia and the United Arab Emirates; in these three systems, missing-value indicators carried about a third of the explained part (Figure 5). In answer to RQ3, confidence, bullying and school-entry skills contributed most. This pattern, with an explained share of about half in both samples, held within classrooms, under alternative references, and across missing-data methods, while school-entry skills and bullying contributed mainly in the lower half of the distribution.

Table 4

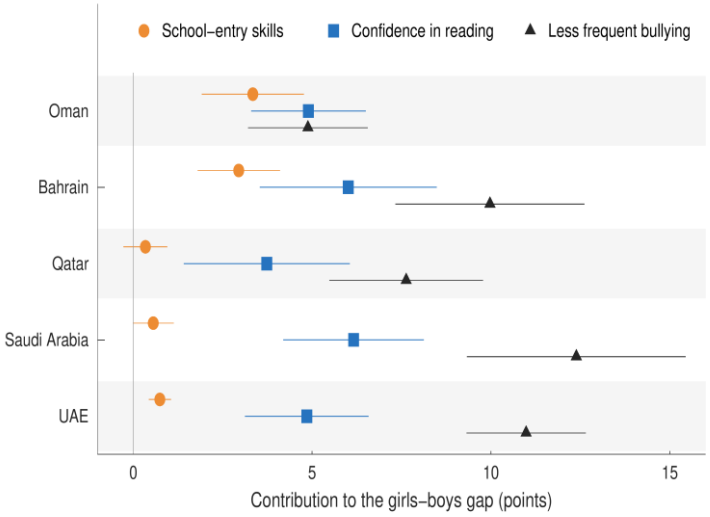
Gender Gap and Its Decomposition in Five GCC Systems (Model C, Full Samples)

System	Mixed classes (%)	Parent data missing (%)	Gap	Explained	Unexplained	Explained (%)
Oman	99.1	3.3	35.55 (3.84)	18.25 (2.17)	17.29 (3.12)	51.4
Bahrain	30.5	6.6	48.93 (4.51)	30.15 (3.04)	18.77 (3.72)	61.6
Qatar	49.1	28.7	16.53 (4.98)	17.91 (3.15)	−1.39 (3.72)	108.7
Saudi Arabia	0.5	26.9	35.39 (6.77)	33.71 (4.05)	1.68 (6.06)	95.3
United Arab Emirates	33.4	49.8	28.67 (5.22)	23.60 (3.19)	5.07 (3.71)	82.3

Note. Score points with standard errors in parentheses; full samples with missing-value indicators in all systems. Oman's values refer to all 5,321 students, not to the complete-data sample of Table 3 (gap 31.64, 49.7%). A share above 100% arises when the unexplained part is negative, that is, when measured characteristics predict a larger gap than the one observed.

Figure 5

Contributions of School-Entry Skills, Confidence and Less Frequent Bullying to the Gap in Five GCC Systems (Model C, full samples), with 95% Confidence Intervals



IN PRESS

5. Discussion

This study examined the gender gap in reading achievement among Omani Grade 4 students who participated in PIRLS 2021 and the extent to which home and school background, early literacy experiences and skills, and students' reading attitudes and school experiences are associated with this gap. The findings show how much of girls' advantage in reading these factors statistically account for and which of them contribute most. By considering these factors together, and by comparing girls and boys taught in the same classrooms, the study contributes to a fuller understanding of gender differences in reading achievement in the Omani educational context. The following discussion addresses each research question in turn, interprets the findings in relation to previous research, considers their educational implications, and concludes with the limitations that should be taken into account when interpreting the results.

In answer to RQ1, Omani girls outperformed boys in reading by about a third of a standard deviation, and the gap was concentrated among weak readers. It mainly reflects a large group of boys with weak reading skills: 44.5% of boys, compared with 31.2% of girls, scored below the Low International Benchmark, and the gap fell from 50.0 points at the 10th percentile to 18.5 points at the 90th. This finding is in line with the long-standing but variable girls' advantage seen internationally (Steinmann et al., 2023), although the Omani gap was about twice the PIRLS 2021 average (Mullis et al., 2023). It also agrees with the large girls' advantages reported for GCC systems (Al-Ubaydli & Alsyoof, 2025) and for South Africa and Morocco (Hofmeyr, 2022; Ibourk & El Aynaoui, 2025).

In answer to RQ2, home and school background accounted for about a tenth of the gap, adding school-entry skills raised this share to about a fifth, and adding students' attitudes and school experiences raised it to about half (49.7% in the complete-data sample and 48.0% to 51.4% in the full sample). The accounted part came mostly from three measures rather than from family or school resources. The near-zero school contribution fits the expectation that girls and boys in mixed classes share school conditions. The small home resources contribution fits the view that family resources vary little between daughters and sons; it is attenuated by the index's low reliability but remains small after correction. The explained share is larger than the roughly one-fifth reported for Morocco (Ibourk & El Aynaoui, 2025), where different covariates were used; this difference suggests that the student measures in PIRLS 2021 account for much of the part accounted for. The large increase when the student block was added agrees with South African evidence that girls' higher reading self-efficacy and enjoyment contribute to their advantage (Hofmeyr, 2022).

About half of the gap remained unaccounted for in both samples and within classes. This residual cannot be read as discrimination or an effect of teacher behaviour, and the analysis identifies no mechanism behind it. It can contain omitted variables, such as leisure reading (Steinmann & Strietholt, 2026), social and behavioural skills (DiPrete & Jennings, 2012) and interest in the texts read (Lepper et al., 2022); measurement error in the included scales, since the corrections for confidence and school-entry skills reduced it from 15.9 to between 14.0 and 15.7 points; differences between girls' and boys' coefficients, which belong to the residual under the pooled reference and were visible for confidence and belonging; and unmeasured aspects of reading engagement, because the engagement scale covers reading lessons only. These sources cannot be separated. The within-classroom result rules out explanations shared by all students of a class but not differential responses of teachers or peers (Aucejo et al., 2022). Across the distribution, the smaller gap among stronger readers reflected mainly a smaller residual, whereas the school-entry and bullying components were concentrated among weaker readers.

In answer to RQ3, confidence, bullying and school-entry skills made the largest contributions. This pattern held within classrooms, under alternative references and across missing-data methods, whereas school-entry skills and bullying contributed mainly in the lower half of the distribution. These three contributions are discussed below in relation to previous studies.

School-entry skills accounted for about a tenth of the gap, including within classes, and absorbed most of the contribution of early literacy activities, as the home literacy model predicts (Hojen et al., 2022; Senechal & LeFevre, 2002). It also agrees with PIRLS evidence that reading with the child before

school predicts Grade 4 achievement (Graham & Mtsweni, 2025) and with the role of pre-primary attendance in the Moroccan gap (Ibourk & El Aynaoui, 2025). Because the measure is a parent report given about three years after school entry, recall error can move the estimate in either direction. Imprecise recall unrelated to the child's sex attenuates the association and understates the contribution, which rose to between 4.1 and 5.0 points under corrections for recall error. Reports coloured by the child's current reading, or more favourable perceptions of daughters' early skills consistent with gendered beliefs about reading (Espinoza & Strasser, 2020), would overstate it; if a quarter or half of the reported sex difference reflected perception, the contribution would fall to about 2.3 or 1.5 points. The direction and approximate size of this contribution, but not its exact value, can therefore be inferred, and the null finding for Türkiye (Coshkun & Goktash, 2026) is a further reason for caution.

Confidence made the largest contribution, consistent with expectancy-value theory and other PIRLS 2021 analyses (Chalishkan & Goktenturk, 2026; Thomure et al., 2025; Yalchin, 2026), but its interpretation is the most uncertain. Under the reciprocal effects model (Marsh & Craven, 2006), weaker reading lowers confidence, so both terms of the contribution (the sex difference in confidence and its coefficient) may partly reflect the gap itself, and the cross-sectional data cannot establish the direction. If confidence were entirely a consequence of reading, the appropriate model would omit it, giving 43.9% rather than 49.7% in the complete-data sample; the reduction is modest because liking reading and bullying take up much of its share. If all concurrent student measures were consequences, the appropriate model would be Model B (18.7%). Model C therefore marks the upper end of what the data attribute to measured characteristics. Measurement error works in the other direction: the scale's modest reliability attenuates its contribution, which rose to between 9.3 and 11.3 points when corrected. However, its negatively worded items, which include comparisons with classmates, are close to judgements of one's own performance. The uncorrected and corrected values therefore mark a plausible range, and the corrected value is not necessarily the better estimate. The 5.5-point contribution is thus an association, not the amount by which the gap would shrink if boys' confidence were raised.

Bullying made the second-largest contribution (4.7 points). Girls reported bullying less often, and less frequent bullying accompanied higher achievement, in line with international evidence (Gimenez et al., 2024; Opoku, 2026), although weaker readers may also be targeted more often. The bullying scale was reliable and invariant across sex. Together, the three largest contributions suggest a gendered pattern that may begin before school and continue in the classroom, consistent with accounts in which reading is defined as feminine and boys' reading self-concept is shaped by adults and peers (Huang et al., 2025; McGeown et al., 2012; Retelsdorf et al., 2015). Because the teaching staff is almost exclusively female, boys meet few male reading models, although the role of this factor cannot be tested here.

The GCC comparison requires caution. Where classes are single-sex, sex is partly confounded with school, so differences between boys' and girls' schools, for example in climate (Basharat & Steinmann, 2025), enter the contributions. Where up to half of the parent data are missing, replacing missing values with means weakens the parent-reported scales, and the missing-value indicators carry a third of the explained part without identifying any factor, so the smaller school-entry contributions there may reflect sparse data. The 108.7% share in Qatar reflects a small gap and an unexplained part that did not differ from zero, rather than a reliable over-prediction of the gap. Scale invariance across countries was also not tested. The comparison therefore shows that the same measures account for the gap to different degrees across Gulf systems (Al-Ubaydli & Alsyoof, 2025), not that the underlying processes differ.

Theoretically, the results show that much of the gap can be divided into an early component, referring to skills before formal instruction, and a larger component associated with concurrent attitudes and school experiences; they also illustrate how nested blocks, within-classroom comparison, and quantile decomposition can be applied within the design of international assessments. Practically, because the contributions are associations rather than causal effects, they identify priorities for further investigation and evaluation, not interventions of established effectiveness. Whether support for shared reading and early literacy at home and in preschool, with outreach to the parents of boys (Al

Maskari et al., 2022), narrows the gap would need longitudinal or experimental evaluation. Because confidence and achievement may reinforce each other, measures aimed at confidence alone cannot be expected to close the gap by the estimated contribution of confidence; whether approaches that build boys' reading skills and experiences of success together, for example with texts matched to their interests (Lepper et al., 2022), reduce the gap remains to be tested. Likewise, whether anti-bullying programmes affect boys' reading is unknown and would need evaluation, given the uncertain direction of the association. Reporting the outcomes of remedial programmes separately for girls and boys and monitoring school-entry skills, confidence and bullying with PIRLS indicators would make such evaluations possible. Whether addressing gendered expectations in teacher education (Retelsdorf et al., 2015) changes boys' reading self-concept could also be examined.

The study has limitations. Because the data are cross-sectional, all contributions are statistical associations, and the student measures may be consequences of achievement. The confidence scale has modest reliability and moderate fit, the home resources index has low internal consistency, and school-entry skills rest on retrospective reports. The complete-data analysis excluded 16.9% of students, disproportionately boys and low achievers; imputation, reweighting, and indicators gave full-sample shares (48.0% to 51.4%) close to the complete-data 49.7%, and a delta adjustment changed the imputed share by up to eight percentage points without changing the leading factors, but larger departures from the missing-at-random assumption cannot be ruled out. The data were collected during the COVID-19 pandemic, and the GCC comparison relies on missing-value indicators where parent data were sparse. Longitudinal studies that follow girls and boys from school entry, direct assessments of early literacy, measures of leisure reading and test engagement, and comparisons of single-sex and mixed Gulf schools would allow stronger inferences.

Acknowledgements

The authors thank the IEA and the TIMSS & PIRLS International Study Center for making the PIRLS 2021 International Database publicly available. During the preparation of this manuscript, the authors used Claude (Anthropic), an AI tool, to assist with drafting and language editing. The authors reviewed, verified, and edited all text, code, results, and references and take full responsibility for the content of the manuscript.

Disclosure Statement

The authors claim no conflict of interest.

Funding

The research did not receive any specific grants from funding agencies.

Data Availability

The PIRLS 2021 International Database analysed in this study is publicly available from the IEA at https://pirls2021.org/data. The supplementary materials, including the supplementary tables and the analysis code, are available to readers upon request from the corresponding author.

Ethics Declaration

The study is a secondary analysis of the anonymised public-use PIRLS 2021 database, collected by national centres under IEA procedures that include informed participation of schools, parents and students. The files contain no direct identifiers. The data were used under the IEA's terms of use, and no further ethical approval was required.

References

Alajmi, M. S., & Al Balushi, S. M. M. A. (2024). Reading comprehension questions in fourth grade Arabic language textbooks in Oman: An analysis from the perspective of Progress in International Reading Literacy Study. *Theory and Practice in Language Studies*, *14*(11), 3355–3364. https://doi.org/10.17507/tpls.1411.04

IN PRESS

Al-Balushi, S. M., Mansour, N., Almehrizi, R. S., Ambusaidi, A. K., & Al-Harthy, I. S. (2022). The association between the gender gap in science achievement and students' perceptions of their own attitudes and capabilities. *EURASIA Journal of Mathematics, Science and Technology Education*, *18*(11), Article em2184. https://doi.org/10.29333/ejmste/12559

Al Khamisi, H., Al Barwani, T., Al Mekhlafi, A., & Osman, M. (2016). EFL reading achievement: Impact of gender and self-efficacy beliefs. *International Journal of Learning, Teaching and Educational Research*, *15*(3), 54–73.

Al Maskari, Z., Al Mawali, F., & Al Zadjali, S. (2022). Oman. In K. A. Reynolds, E. Wry, I. V. S. Mullis, & M. von Davier (Eds.), *PIRLS 2021 encyclopedia: Education policy and curriculum in reading*. TIMSS & PIRLS International Study Center, Boston College. https://pirls2021.org/wp-content/uploads/2022/10/Oman.pdf

Al-Ubaydli, O., & Alsyoof, M. (2025). An analysis of the gender gap in GCC educational attainment. *Economy of the Middle East*, *2*(1).

Aucejo, E. M., Fruehwirth, J. C., Kelly, S., & Mozenter, Z. (2022). Teachers and the gender gap in reading achievement. *Journal of Human Capital*, *16*(3), 372–403. https://doi.org/10.1086/719731

Bailey, P., Webb, B., Kelley, C., Nguyen, T., & Huo, H. (2024). *WeMix: Weighted mixed-effects models using multilevel pseudo maximum likelihood estimation* (Version 4.0.4) [Computer software]. https://CRAN.R-project.org/package=WeMix

Basharat, S., & Steinmann, I. (2025). Divided desks, divided school climate conditions? Investigating the implications of single-sex schooling. *Equity in Education & Society*. https://doi.org/10.1177/27526461251348603

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, *8*(4), 436–455.

Chalishkan, E. A., & Goktenturk, T. (2026). Ecological dynamics of reading performance: A structural equation model with PIRLS 2021 Türkiye data. *Participatory Educational Research*, *13*(4), 1–22. https://doi.org/10.17275/per.26.46.13.4

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Claes, R., Laga, J., Denies, K., Bleukx, N., Dockx, J., Van Keer, H., & Aesaert, K. (2024). The interplay between students' home literacy environment, reading attitudes and comprehension: A serial mediation analysis using PIRLS 2021-data. *Large-scale Assessments in Education*, *12*, Article 43. https://doi.org/10.1186/s40536-024-00233-8

Coshkun, B., & Goktash, S. A. (2026). The role of home and classroom factors in shaping reading achievement: A multilevel study based on PIRLS 2021 Türkiye data. *Studies in Educational Evaluation*, *89*, Article 101598. https://doi.org/10.1016/j.stueduc.2026.101598

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

Espinoza, A. M., & Strasser, K. (2020). Is reading a feminine domain? The role of gender identity and stereotypes in reading motivation in Chile. *Social Psychology of Education*, *23*(4), 861–890. https://doi.org/10.1007/s11218-020-09571-1

Espinoza Catalan, A. M., Strasser Salinas, K., & Carvacho García, H. S. (2025). Effects of sex, gender stereotypes and gender identity in male and female students' reading motivation. *Psykhe*, *34*, 1–23. https://doi.org/10.7764/psykhe.2023.58757

Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, *77*(3), 953–973. https://doi.org/10.3982/ECTA6822

Fishbein, B., Yin, L., & Foy, P. (2024). *PIRLS 2021 user guide for the international database* (2nd ed.). TIMSS & PIRLS International Study Center, Boston College. https://pirls2021.org/data

Fortin, N. M. (2008). The gender wage gap among young adults in the United States: The importance of money versus people. *Journal of Human Resources*, *43*(4), 884–918. https://doi.org/10.3368/jhr.43.4.884

Foy, P., & Almaskut, A. (2023). Estimating standard errors in the PIRLS 2021 results. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and procedures: PIRLS 2021 technical report* (pp. 13.1–13.31). TIMSS & PIRLS International Study Center, Boston College. https://doi.org/10.6017/lse.tpisc.tr2103.kb1576

Gelbach, J. B. (2016). When do covariates matter? And which ones, and how much? *Journal of Labor Economics*, *34*(2), 509–543. https://doi.org/10.1086/683668

Gimenez, G., Mediavilla, M., Giuliodori, D., & Rusteholz, G. C. (2024). Bullying at school and students' learning outcomes: International perspective and gender analysis. *Journal of Interpersonal Violence*, *39*(11–12), 2733–2760. https://doi.org/10.1177/08862605231222457

Graham, M. A., & Mtsweni, M. U. (2025). Parental involvement predicts Grade 4 learners' reading literacy: An analysis of PIRLS data for students in Mpumalanga, South Africa. *Educational Review*, *77*(6), 1907–1929. https://doi.org/10.1080/00131911.2024.2379416

Hofmeyr, H. (2022). Why do girls do better? Unpacking South Africa's gender gap in PIRLS and TIMSS. *International Journal of Educational Development*, *94*, Article 102648. https://doi.org/10.1016/j.ijedudev.2022.102648

Hojen, A., Schmidt, A. S. M., Møller, I. S., & Flansmose, L. (2022). Unequal home literacy environments between preschool-age boys and girls predict unequal language and preliteracy outcomes. *Acta Psychologica*, *230*, Article 103716. https://doi.org/10.1016/j.actpsy.2022.103716

Huang, M., Yi, H., & Rozelle, S. (2025). On the origins of gender gaps in education: Stereotype as a self-fulfilling prophecy. *Journal of Population Economics*, *38*, Article 45. https://doi.org/10.1007/s00148-025-01102-6

Ibourk, A., & El Aynaoui, K. (2025). From past to present: Tracking gender performance in Moroccan education across PIRLS and TIMSS iterations. *Cogent Education*, *12*(1), Article 2551395. https://doi.org/10.1080/2331186X.2025.2551395

Jann, B. (2008). The Blinder–Oaxaca decomposition for linear regression models. *The Stata Journal*, *8*(4), 453–479. https://doi.org/10.1177/1536867X0800800401

Jones, M. P. (1996). Indicator and stratification methods for missing explanatory variables in multiple linear regression. *Journal of the American Statistical Association*, *91*(433), 222–230. https://doi.org/10.1080/01621459.1996.10476680

Kazandjian, N., Harandian, K., Routhier-Guilmette, S., Dufour, M.-M., Archambault, I., & Pagani, L. S. (2025). From storytime to success: Prospective longitudinal associations between toddler literacy enrichment and long-term student engagement in a millennial birth cohort of boys and girls. *Journal of Intelligence*, *13*(6), Article 66. https://doi.org/10.3390/jintelligence13060066

Kennedy, A. I., & Strietholt, R. (2023). School closure policies and student reading achievement: Evidence across countries. *Educational Assessment, Evaluation and Accountability*, *35*(4), 475–501. https://doi.org/10.1007/s11092-023-09415-4

Lepper, C., Stang-Rabrig, J., & McElvany, N. (2022). Gender differences in reading: Examining text-based interest in relation to text characteristics and reading comprehension. *Learning and Instruction*, *82*, Article 101680. https://doi.org/10.1016/j.learninstruc.2022.101680

Li, Y., Jia, K., Myat Kyaw, H. M., Li, H., & Yan, M. (2024). The unique contribution of reading motivation to reading comprehension increases from Grades 2 to 4 in Chinese children. *Journal of Research in Reading*, *47*(4), 557–579. https://doi.org/10.1111/1467-9817.12466

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, *1*(2), 133–163. https://doi.org/10.1111/j.1745-6916.2006.00010.x

McGeown, S., Goodwin, H., Henderson, N., & Wright, P. (2012). Gender differences in reading motivation: Does sex or gender identity provide a better account? *Journal of Research in Reading*, *35*(3), 328–336. https://doi.org/10.1111/j.1467-9817.2010.01481.x

Mullis, I. V. S., von Davier, M., Foy, P., Fishbein, B., Reynolds, K. A., & Wry, E. (2023). *PIRLS 2021 international results in reading*. TIMSS & PIRLS International Study Center, Boston College. https://doi.org/10.6017/lse.tpisc.tr2103.kb5342

Neumark, D. (1988). Employers' discriminatory behavior and the estimation of wage discrimination. *Journal of Human Resources*, *23*(3), 279–295.

Oaxaca, R. (1973). Male–female wage differentials in urban labor markets. *International Economic Review*, *14*(3), 693–709.

Opoku, M. P. (2026). Bullying in early school years: Bullying as a predictor of Grade 4 students' attitudes and achievement in mathematics and science. *International Journal of Bullying Prevention*.https://doi.org/10.1007/s42380-026-00358-0

Osman, M. E., Al Barwani, T. A., & Almekhlafi, A. (2015). The male dilemma: Patterns of gender disparity in academic performance in Oman. *Global Journal for Research Analysis*, *4*(10), 39–42.

Rabe-Hesketh, S., & Skrondal, A. (2006). Multilevel modelling of complex survey data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *169*(4), 805–827. https://doi.org/10.1111/j.1467-985X.2006.00426.x

Rakhlin, N. V., Li, N., Aljughaiman, A., & Grigorenko, E. L. (2025). "Speech is golden": The importance of colloquial Arabic for reading standard Arabic for beginning readers. *Journal of Speech, Language, and Hearing Research*, *68*(3S), 1441–1467. https://doi.org/10.1044/2024_JSLHR-23-00522

R Core Team. (2025). *R: A language and environment for statistical computing* (Version 4.5.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org

Retelsdorf, J., Schwartz, K., & Asbrock, F. (2015). "Michael can't read!": Teachers' gender stereotypes and boys' reading self-concept. *Journal of Educational Psychology*, *107*(1), 186–194. https://doi.org/10.1037/a0037107

Ridge, N. (2014). *Education and the reverse gender divide in the Gulf states: Embracing the global, ignoring the local*. Teachers College Press.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Saiegh-Haddad, E. (2003). Linguistic distance and initial reading acquisition: The case of Arabic diglossia. *Applied Psycholinguistics*, *24*(3), 431–451. https://doi.org/10.1017/S0142716403000225

Sakellariou, C. (2025). Reciprocal longitudinal effects between sense of school belonging and academic achievement: Quasi-experimental estimates using United States primary school data. *Frontiers in Psychology*, *15*, Article 1478320. https://doi.org/10.3389/fpsyg.2024.1478320

Senechal, M., & LeFevre, J. A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development*, *73*(2), 445–460. https://doi.org/10.1111/1467-8624.00417

Steinmann, I., & Strietholt, R. (2026). Reading more, achieving more? Long-term gender gaps in reading behavior and reading achievement. *Studies in Educational Evaluation*, *88*, Article 101581. https://doi.org/10.1016/j.stueduc.2026.101581

Steinmann, I., Strietholt, R., & Rosén, M. (2023). International reading gaps between boys and girls, 1970–2016. *Comparative Education Review*, *67*(2), 298–330. https://doi.org/10.1086/724089

Thomure, H. T., Oo, C. Z., & Khine, M. S. (2025). Reading achievements among students in the United Arab Emirates: A multilevel analysis of PIRLS 2021. *International Journal of Educational Research Open*, *8*, Article 100421. https://doi.org/10.1016/j.ijedro.2024.100421

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Yalchin, G. (2026). School and student variables of reading achievement: A multilevel mediation model. *Participatory Educational Research*, *13*(3), 239–259. https://doi.org/10.17275/per.26.43.13.3

Yin, L., & Reynolds, K. A. (2023). Creating and interpreting the PIRLS 2021 context questionnaire scales. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and procedures: PIRLS 2021 technical report* (pp. 15.1–15.161). TIMSS & PIRLS International Study Center, Boston College. https://doi.org/10.6017/lse.tpisc.tr2103.kb6994

IN PRESS