Iranian Journal of Comparative Education, 2026, 9(4), 4442-4460

Research Article **DOI: 10.22034/IJCE.2026.576268.1706** http://journal.cesir.ir

A Comparative Study of Trends in Reading Literacy among Iranian Students: Evidence from PIRLS 2011–2021

Maryam Mohsenpour [1]
Azam Moghadam [2] (Corresponding author)

ARTICLE INFO

Received: 14 February 2026
Revised: 10 May 2026
Accepted: 05 June 2026
Online: 27 November 2026

K E Y W O R D S

Large-scale Assessments
Reading Literacy
PIRLS
Trend

ABSTRACT

This Iranian fourth-grade students' reading literacy scores showed a declining trend across three PIRLS cycles: 457 in 2011, 428 in 2016, and 413 in 2021. This trend study applied multilevel modeling to compare student-level (Students Like Reading, Students Confident in Reading) and school-level (School Emphasis on Academic Success) predictors across the three cycles. Data were analyzed using student- and teacher-level weights via the EdSurvey package, and restricted maximum likelihood estimation was used for two-level random-intercept models. Results showed that based on full model the predictive power of "Students Confident in Reading" increased significantly from 2011 ($B = 10.81$, $p < .001$) to 2016 ($B = 16.67$, $p < .001$), then slightly decreased but remained strong in 2021 ($B = 15.44$, $p < .001$). In contrast, "Students Like Reading" showed a decreasing trend, dropping from $B = 5.94$ ($p < .001$) in 2011 to $B = 4.18$ ($p < .05$) in 2016, and became non-significant in 2021 ($B = 1.23$, $p > .05$). "School Emphasis on Academic Success" had modest but significant effects in all three cycles ($B = 6.11$ in 2011, 4.41 in 2016, and 7.05 in 2021; all $p < .05$). Between-school variance remained high (ICC ≈ 0.38–0.40) across the decade, which was nearly double the PIRLS 2021 international average (0.21). It is recommended that educational policymakers implement interventions to enhance "Student Confidence in Reading" and pay attention to reducing between-school inequities.

[1] Associate Professor, Department of Educational Psychology, Faculty of Education and Psychology, Alzahra University, Tehran, Iran. E-mail: m.mohsenpour@ut.ac.ir
[2] Ph.D. Department of Assessment & Measurement, Faculty of Psychology and Educational Sciences, Allameh Tabataba'i University, Tehran, Iran. E-mail: azam.moghadam@gmail.com

1. Introduction

Studies indicate that students who develop strong reading skills in the early grades tend to maintain their proficiency in later years, whereas those who have difficulty reading in first grade are likely to continue struggling by higher grades (Wagner, 2017; Park et al., 2015; Fuchs et al., 2012). Furthermore, the reading abilities a child develops early in life are vital to their future academic success or struggles, as literacy is deeply connected to all other subjects in school. Acquiring reading skills and fostering a love of reading significantly influence a child's personal growth and attitude formation (Ozola & Geske, 2019). Moreover, proficient reading skills are essential for both academic achievement and social engagement (Min & Seo 2024; Thumore et al., 2025; Lopez et al., 2022; Schiefele et al., 2012).

Conscious of the significance of reading in contemporary societies, the International Association for the Evaluation of Educational Achievement (IEA)—a collaboration of governmental and non-governmental institutes and agencies dedicated to educational research—has been conducting the Progress in International Reading Literacy Study (PIRLS). PIRLS is a global assessment that compares students' reading performance across countries. Conducted every five years since 2001, it evaluates the reading skills of fourth-grade students and has gathered over two decades of trend data from 2001 to 2021 (Mullis et al., 2020). Currently, the PIRLS definition of reading literacy is as follows:

> *Reading literacy is the ability to understand and use those written language forms required by society and/or valued by the individual. Readers can construct meaning from texts in a variety of forms. They read to learn, to participate in communities of readers in school and everyday life, and for enjoyment (*Mullis & Martin, 2019, p. 6*).*

In general, PIRLS has two main purposes for reading including literary experience and acquire and use information (Mullis et al., 2023). Each student participating in PIRLS is evaluated on their proficiency in comprehending literary fiction as well as concise informational texts that cover basic topics in science, geography, and history. Furthermore, every test item in the PIRLS assessments is designed to assess one of four key reading comprehension skills. These skills, arranged in order of increasing complexity, include retrieving explicitly stated information, making straightforward inferences, interpreting and integrating ideas across texts, and evaluating texts, which involves identifying the author's viewpoints or the effects of particular language choices (Wry & Mullis, 2023). According to PIRLS theoretical framework, throughout the reading process—

before, during, and after—individuals draw upon a range of linguistic abilities, cognitive and metacognitive strategies, and prior knowledge to derive meaning from text. Furthermore, the situational context of reading can facilitate comprehension by fostering engagement and motivation; however, it may also impose specific cognitive demands that hinder the construction of meaning (Mullis & Martin, 2019). So, in addition to measuring reading achievement, PIRLS collects extensive information on the national, home, school, and classroom environments where students learn to read and read to learn. This contextual data offers an international perspective on the educational factors influencing reading outcomes (Fishbein, Yin & Foy, 2023). **Therefore,** the PIRLS datasets provide valuable resources that can be used to explore numerous research questions across education, psychology, and various social science fields (Stiff et al., 2023).

Iran participated in PIRLS for the first time in 2001 and has since take part in five assessment cycles: 2001 (mean score: 414), 2006 (421), 2011 (457), 2016 (428), and 2021 (413). The most notable improvement and the highest recorded performance occurred in 2011, with a mean score of 457. However, in 2016, the score declined by 29 points, bringing the national average down to 428**.** In the latest PIRLS cycle (2021), Iran was ranked among the countries scoring below the international centerpoint of the scale (500), with a national mean reading achievement score of 413 (Mullis et al., 2023).

PIRLS employs a rigorous multistage sampling design and a hierarchical structure suitable for multilevel analysis. The dataset includes plausible values and a wide range of contextual variables, adheres to strict international standards, and is publicly accessible for secondary research. Despite the considerable potential of PIRLS data, diverse multilevel studies utilizing secondary analysis of these data have not yet been conducted in Iran. Therefore, given the observed declining trend in students' performance in PIRLS assessments, the present study aims to identify and examine factors influencing students' reading literacy across the 2011, 2016, and 2021 cycles using a multilevel approach. A notable feature of this study is that the selection of variables is based on four criteria. First, their potential for intervention and modifiability within the educational system such as confident in reading and teacher job satisfaction. Home-based contextual variables such as "home resources" and "early literacy tasks" that lie outside the direct control of the educational system, and therefore cannot be manipulated or improved by schools or education policy system**.** Second, the variables were drawn from scales developed by PIRLS. Third, only scales that were repeated across all three cycles were selected. Fourth, Variables with a correlation greater than 0.1 with reading achievement in each cycle were included . Based on Cohen's (1988) guidelines for effect sizes in correlation analysis, we adopted a threshold of $r \geq 0.10$ for retaining

predictor variables in the multilevel model. According to Cohen, r = 0.10 represents the smallest effect size that is still practically meaningful in social and behavioral sciences (referred to as a 'small' effect). Finally, the variables selected as predictors of reading achievement included “students like reading” and “students confident in reading” at the student level, and “school emphasis on academic success” at the school level.

2. Literature Review

According to Marôco (2021), the PIRLS assessment framework is strongly aligned with the Component Model of Reading (Aaron et al., 2008), as it incorporates a broad spectrum of potential factors that explain reading literacy. These encompass student-related variables from both the cognitive and psychological domains, as well as environmental factors related to families, teachers, and schools. Based on the Component Model of Reading (Aaron et al., 2008), a child's reading literacy is shaped by a combination of cognitive, psychological, and environmental factors. This includes not only skills like word recognition and fluency but also elements such as self-motivation, teacher expectations, home environment, and school resources.

A large body of literature provides evidence that differences in students' reading achievement are associated with student-level factors (Xiao, 2023; Lopes et al., 2021; Yang et al., 2018; Cheung e al., 2017; Park, 2011). Among those variables, motivational factors play a major role. According to Schiefele et al. (2012), a significant body of research has shown that students' motivation influences both how they learn and what they achieve, going beyond just their cognitive traits like intelligence or prior knowledge. As stated by Mullis et al. (2023), among motivational variables, confidence in reading (as a dimension of reading attitude) has been identified as a strong indicator of reading success. Confidence in reading which in the PIRLS research literature, sometimes referred to as self-concept or self-efficacy, is described as the overall collection of an individual’s beliefs about their own abilities.

For instance, Xiao (2023) examined reading motivation and reading self-concept and their relationships with reading behavior and performance using Hong Kong data from PIRLS 2016. The study highlighted the indirect effects of motivation on reading achievement through reading amount. The results showed that reading attitude and self-concept influence achievement directly and indirectly via reading amount. Furthermore, Lopes et al. (2022), by analyzing data from 4,118 fourth graders in the IEA PIRLS 2016 using hierarchical linear modeling, investigated whether student factors predict reading performance better than school factors and identified key predictors at both levels. They found that students’ reading confidence and home learning

resources are the strongest student-level predictors, while school emphasis on academic success is the top predictor at the school level. Also, Yang et al. (2018) investigated how motivation and home environment factors influence the reading literacy of 4th-grade students in Abu Dhabi using data from the 2011 PIRLS. The study found that reading self-efficacy (self-confidence in reading) was the strongest predictor of reading success. Several home-related factors also had notable positive effects. Surprisingly, though, two commonly emphasized elements—intrinsic motivation and parental involvement in learning—were found to have no effect or even a negative impact on students' reading performance.

An additional motivational variable which has been investigated in secondary analysis of PIRLS data, is like reading, which is usually described in the literature as reading for pleasure or attitude towards reading (Lopes et al., 2022). According to Mullis et al. (2023) in PIRLS 2021, there was a small correlation between enjoying reading and reading performance in fourth-grade students. However, the results of studies in this regard are **inconsistent**. For instance, Tse & Xiao (2014), in their research on PIRLS 2011 analyzed data from 3,875 fourth-grade students in Hong Kong to explore how reading-related affective factors and home and school environments influence students' reading proficiency levels. Using multilevel regression, the research found that positive reading attitude, like reading, self-confidence in reading, and lower levels of peer and school bullying, as well as higher school socioeconomic status, increase the likelihood of high reading proficiency. In contrast, low reading proficiency was mainly linked to lower reading self-concept and experiences of peer bullying. What is more, in a study conducted by Taha Thomure et al. (2025), the authors examined factors influencing reading achievement among students in the UAE using PIRLS 2021 data. Multilevel modeling revealed that at the student level, gender, socioeconomic status, reading confidence, and reading engagement significantly affected reading achievement, while simply liking reading did not. At the school level, only two factors—teachers sending students to the library and overall school characteristics—had significant impacts.

Another factor influencing students' reading performance is school-level characteristics (Goddard et al., 2000; Mullis et al., 2017; Vlachou & Stefanou, 2017). One of these characteristics is teachers' perception of school emphasis on academic success, which, based on Mullis et al. (2021), is a dimension of school climate and consists of series of teacher, parent, and student attitudes and actions that can contribute to academic success. According to Nilsen & Gustafsson (2014, 2016), a stronger school's emphasis on academic success is positively related to higher academic achievement. Moreover, Mullis et al. (2023)–concluded that students attending in schools that placed a stronger emphasis on academic success performed better in reading than those in schools

with a weaker emphasis on academic success. In another study, Pongsophon (2024) analyzed student- and school-level factors affecting reading performance in five top-performing countries from PIRLS 2021 using multilevel analysis. Key student predictors included gender, digital self-efficacy, and school belonging, with their impacts differing by country. At the school level, the number of computers, principal experience, and academic emphasis significantly affected reading scores. The combined model explains a large part of the reading score variance, particularly in Hong Kong and Russia. In addition, Thomure et al. (2025) examined factors affecting students' reading achievement in the UAE using PIRLS 2021 data. Their study found that most student-level factors—except for liking reading—significantly influenced reading performance. At the school level, only teachers encouraging library visits and certain school characteristics (emphasis on academic success and school discipline) had a notable impact. Additionally, the Costa & Araújo (2017) study explored the relationship between reading achievement and both student and school characteristics in countries with varying fourth-grade reading performance levels. It used data from the 2011 PIRLS for Denmark, Sweden, and France. The findings showed that only in France, did students tend to perform better academically when they attended schools that placed a stronger emphasis on academic success.

Several studies have applied multilevel modeling to Iranian PIRLS data, yet their findings remain inconsistent. Using PIRLS 2006 data, Rezaei et al. (2015) found that student self-concept and attitude had no significant effect on reading achievement at the student level, while the amount of study had a positive effect. In contrast, Zare and Naqsh (2016)—also using PIRLS 2006 data—reported that both self-concept and attitude significantly predicted reading achievement. More recently, Mohammadnia & Mohsenpour (2025), using PIRLS 2021 data, showed that self-confidence in reading significantly predicted reading literacy, whereas attitude towards reading and teachers' perception of school emphasis on academic success did not.

These contradictions may arise from differences in model specification, sample characteristics, or—critically—changes across PIRLS cycles in the Iranian educational context. However, no previous study has systematically compared the relationships between student-level factors (self-concept, attitude, self-confidence) and reading literacy across multiple PIRLS cycles in Iran. In particular, the period from 2011 to 2021—covering three successive PIRLS assessments (2011, 2016, and 2021)—remains unexamined from a trend perspective. To address this gap, the present study investigates how the predictive power of student and school level variables has changed across these three cycles. By doing so, this study aims to determine whether the observed

inconsistencies in previous research reflect actual changes in the Iranian educational system over time or are merely artifacts of model specification.

In Iran, where a national assessment system for primary education does not exist, results from international large-scale assessments are frequently employed as key indicators of students' academic achievement. As noted earlier, few studies have investigated factors associated with the reading achievement of Iranian students using PIRLS data. Therefore, this study is a secondary analysis dedicated to identifying the most effective student and school-level variables for predicting the reading achievement of 4th grade Iranian students across three assessment cycles.

This study considers three indicators related to reading achievement, which were developed by the IEA Data Center (as measured by the PIRLS index). At the student level, the indicators are students like reading, students confident in reading and at the school level, the indicator is school emphasis on academic success. This study is designed to explore the following research questions:

- Q1- How has student and school factors changed from PIRLS 2011 to 2021 in Iran?
- Q2-Which student and school factors predict reading literacy among Iranian students over the period from 2011 to 2021?

3. Research Method

This study specifically investigates the average reading literacy performance of students, the variation in achievement both within and between schools, and the extent to which scores vary as background variables across assessment cycles. The findings provide important evidence for identifying factors that can improve reading literacy in Iran. This paper uses data on fourth grade students' reading achievement from three cycles of PIRLS in Iran from (PIRLS 2011, 2016, and 2021). In all of these cycles, data were collected through paper-based assessments. In addition, Contextual data were also gathered through students and teachers' questionnaires.

Table 1. Participating sample sizes in PIRLS 2011, 2016 and 2021

	PIRLS 2011	PIRLS 2016	PIRLS 2021
Number of students (unweighted)	5758	4385	5962
Number of students (weighted)	1043012	1200537	1341065
Number of schools	244	271	218

Note: For 2016, 32 schools with non-constant teacher weights were excluded from the multilevel analysis, resulting in a final analytic sample of 239 schools.

PIRLS employs a two-stage random sampling design, first selecting schools and then intact classes within those schools. Sampling whole classes allows the study to capture students' classroom-based instructional experiences while minimizing disruption to school activities. (Almaskut, LaRoche, & Foy, 2023). Table 1 presents the achieved sample sizes for each cycle of PIRLS.

The PIRLS assessment aims to offer a comprehensive overview of students' reading literacy performance in each participating country (Mullis et al., 2020). In addition, it gathers information on the home, school, and classroom environments where students develop reading skills. A number of the items from the PIRLS contextual background instruments were consolidated into composite scales, with each scale representing a unitary latent variable linked to reading performance. Several of these scales have maintained relevance across multiple PIRLS assessment cycles. They were all developed using item response theory (IRT), with the Rasch partial credit model (Masters, 1982; Masters & Wright, 1997) employed to construct the scales, ensuring measurement of the underlying latent constructs associated with reading achievement (Von Davier et al. 2023). This study focuses on scales which were incorporated into three cycles of 2011 -2021, comprising two student- level variable (students like reading, students confident in reading) and one school-level variables (school emphasis on academic success) along with reading achievement scores.

Reading literacy reading comprehension in PIRLS is assessed through a matrix that crosses two reading purposes—literary and informational—by four comprehension strategies (retrieval, inferencing, integrating, and evaluation). In the PIRLS, students initially engage with reading of narrative texts that tell a story or informational texts that tell students about the world around them and each student is randomly assigned a test booklet including two passages and their items. Also, the assessment included 18 passages in 2021, compared to 10 passages in both 2016 and 2011. In PIRLS, reading literacy scores are generated through five plausible values for each student by item response theory (IRT) scaling methods (Bezirhan et al., 2023). In this study we used all plausible value in our analysis.

Students like reading: Students were scored on the Students Like Reading scale according to their degree of agreement with some statements with a four-point Likert response from agree a lot, agree a little, disagree a little and disagree a lot. Cronbach's alpha coefficients for the scale were reported as 0.74 with 8 items in 2021, 0.71 with 10 items in 2016, and 0.52 with 6 items in 2011. Our examination of PIRLS 2011 Cronbach's alpha values for the 'Students Like Reading' scale showed that coefficients for several countries (Azerbaijan, Oman, Qatar, Saudi Arabia, Iran, and Indonesia) ranged from 0.38 to 0.60, below the 0.70 threshold. As secondary analysts, we remained

faithful to the original PIRLS scale without modifications. So, findings from the 2011 cycle should be interpreted with some caution

Students confident in reading**:** Students confident in reading scale was created based on students' degree of agreement with some statement with a four-point Likert response from agree a lot, agree a little, disagree a little and disagree a lot. In 2021 and 2016, the scale comprised 6 items with a Cronbach's alpha coefficient of 0.70, while in 2011 it included 7 items with an alpha of 0.54."

School emphasis on academic Success**:** Students were scored according to their teachers' responses characterizing some aspects on the school climate with a five-point Likert including very high, high, medium, Low and Very low. In 2021 and 2016, the scale comprised 12 items with Cronbach's alpha values of 0.89 and 0.91, respectively, while in 2011 it included 5 items with an alpha of 0.71.

To examine reading literacy, we conducted independent t-tests, correlation analyses, and multilevel linear regression. Trends in average performance and score dispersion were analyzed across cycles. Bivariate analyses used Hedges' \(g\) as the effect size for pairwise comparisons, following PIRLS conventions (Mullis et al., 2017; von Davier et al., 2023).

Multilevel linear regression examined student-level (like reading, confidence in reading) and school-level (school emphasis on academic success) predictors, as well as the intraclass correlation coefficient (ICC). The ICC quantifies the proportion of total variance attributable to between-school differences, often interpreted as systemic inequality (Martin et al., 2017). An ICC > 0.05 justifies multilevel modeling to account for clustered data (Hox et al., 2018; Snijders & Bosker, 2012). Conditional \(R^2\) estimated the variance explained by predictors; model parameters and explained variance were compared across the three PIRLS cycles (2011, 2016 and 2021).

Two-level random-intercept models were estimated using the EdSurvey package (v4.0-1; Bailey et al., 2023) in R (v4.4.3; R Core Team, 2024). Student-level weights (`TOTWGT`) served as base weights; when school/teacher predictors were included, teacher-level weights (`TCHWGT`) were automatically incorporated. Restricted maximum likelihood (REML) was used. Missing data were low (<5% on all variables), ensuring unbiased estimates (Enders, 2010; von Davier et al., 2023).

PIRLS uses a two-stage stratified cluster design (schools, then classes), leading to unequal selection probabilities. To ensure generalizability to the Iranian fourth-grade population, sampling weights were applied in all analyses. Univariate and bivariate analyses used `TOTWGT` and replicate weights to compute correct standard errors (Foy & LaRoche, 2020; LaRoche & Foy, 2017).

For multilevel models, EdSurvey automatically handled the five plausible values by estimating the model separately for each plausible value and combining coefficients and standard errors using Rubin's rules (Rubin, 1987), thereby accounting for imputation uncertainty.

4. Findings

Q1- How has student and school factors changed from PIRLS 2011 to 2021 in Iran?

Table 2 presents the means and standard deviations of all indices for Iranian fourth-grade students across the three PIRLS cycles.

Table 2 Means and standard deviations of all variables across PIRLS cycles

variable	PIRLS 2011	PIRLS 2016	PIRLS 2021
Students Like Reading	10.55 (1.71)	11.23 (1.81)*	11.17 (1.96)
Students Confident in Reading	10.22 (1.99)	10.37 (2.09)	10.24 (2.03)
School Emphasis on Academic Success	10.42 (1.88)	10.06 (2.15)	10.52 (2.21)

*Values are both statically and practically different from the corresponding values in the previous cycle Statistically significant and practically meaningful change from the previous cycle (Hedges' g = .38, small effect).

As shown in Table 2, mean scores for Students Like Reading, Students Confident in Reading, and School Emphasis on Academic Success exhibited only minor variations across PIRLS 2011, 2016, and 2021. Independent-samples t-tests were conducted to examine mean differences between cycles. It is worth mentioning missing data analysis revealed that the rate of missing values for all predictor variables was below 4% across all three PIRLS cycles. The results of the independent-samples t-tests indicated that observed changes in some of the variables were statistically significant. However, the calculation of Hedges' g effect sizes revealed that with the exception of the increase in 'Students Like Reading' from 2011 to 2016 ($g = 0.38$, representing a small effect), the effect sizes of the other significant changes were negligible from a practical standpoint ($|g| < 0.20$).

Q2-Which student and school factors predict reading literacy among Iranian students over the period from 2011 to 2021?

Table 3 presents the average correlations between scale scores and reading achievement.

Table 3: Pearson correlation matrix among all variables and reading literacy

Variable	Reading literacy 2011	Reading literacy 2016	Reading literacy 2021
Students Like Reading	0.280***	0.209***	0.158***
Students Confident in Reading	0.331***	0.435***	0.418***
School Emphasis on Academic Success	0.196***	0.096	0.152***

Note: *** $p < 0.001$; For School Emphasis in 2016: $p = 0.086$ (not significant)

As shown in Table 3, the correlations between the predictors and reading achievement were generally small to moderate based on Cohen's (1988) benchmarks for correlation coefficients. Specifically, Students Confident in Reading showed moderate correlations across all three cycles ($r = 0.331$ in 2011, $r = 0.435$ in 2016, and $r = 0.418$ in 2021). Students Like Reading showed small correlations ($r = 0.280$, $r = 0.209$, and $r = 0.158$, respectively), while School Emphasis on Academic Success showed small correlations in 2011 ($r = 0.196$) and 2021 ($r = 0.152$) and a non-significant correlation in 2016 ($r = 0.096$, $p = 0.086$). When comparing the magnitude of changes in correlations across cycles using Cohen's q (where $|q| \approx 0.10$ = small, $|q| \approx 0.30$ = medium, and $|q| \approx 0.50$ = large), the change in Students Like Reading from 2011 to 2021 was small ($q = 0.128$). Similarly, the changes in Students Confident in Reading ($q = -0.122$ from 2011 to 2016 and $q = -0.102$ from 2011 to 2021) and School Emphasis on Academic Success (q range = −0.057 to 0.103) were small in magnitude. Thus, while the strength of the correlations themselves ranged from small to moderate, the changes in the correlations across cycles were generally small in magnitude.

Multilevel linear regression

To examine the contribution of student- and school-level factors to reading achievement, two-level random-intercept models were estimated separately for each PIRLS cycle. For each cycle, three nested models were fitted: (1) a null (unconditional) model containing only the intercept; (2) a Level-1 model adding the two student-level predictors (Students Like Reading and Students Confident in Reading); and (3) a full model additionally including the school-level predictor (School Emphasis on Academic Success).

Table 4 Percent of between-school variance in reading literacy score

	PIRLS 2011	PIRLS 2016	PIRLS 2021
ICC	39.74%	38.97%	40.60%

The null models revealed substantial clustering of reading achievement at the school level across all three cycles (Table 4). The intraclass correlation coefficient (ICC) ranged from 0.3897

(38.97%) in 2016 to 0.4060 (40.60%) in 2021, indicating that between 38.97% and 40.60% of the total variance in reading scores was attributable to differences between schools. These values are considerably higher than typical ICCs in many high-income countries (often below 0.20–0.25) and rank among the higher figures observed in PIRLS-participating nations (Mullis et al., 2023). The consistently large between-school variance highlights persistent systemic inequalities in the Iranian primary education system and strongly justifies the use of multilevel modelling. Despite minor fluctuations, the proportion of variance explained by school membership remained relatively stable over the decade, suggesting no substantial reduction in educational segregation during this period.

The ICC ranged from 0.390 to 0.406 across cycles, indicating that 39–41% of reading variance was due to between‑school differences. This is nearly twice the PIRLS 2021 average of 0.21 (Mullis et al., 2023), placing Iran among the highest in the assessment. Such high between‑school variance means students' reading achievement strongly depends on which school they attend—a pattern typical of systems with pronounced segregation, unequal resources, or regional disparities. In Iran, this likely reflects urban‑rural quality gaps, socioeconomic stratification across districts, and unequal access to qualified teachers and materials. This persistent variation, largely unchanged from 2011 to 2021, points to structural inequalities that have not been alleviated over the decade.

For the multilevel analyses using the mixed.sdf function in the EdSurvey package (R), consistent level-2 weights within schools were required. In the PIRLS 2011 and PIRLS 2021 Iran datasets, all schools (244 in 2011 and 218 in 2021) had constant teacher weights (tchwgt) within each school ID (idschool), and no exclusions were necessary; the models converged successfully without any weight-related errors. However, in the PIRLS 2016 Iran data, 32 schools (11.8% of the initial 271 schools) exhibited varying teacher weights within the same school ID (i.e., non-zero standard deviation of tchwgt or multiple distinct values), which caused the estimation procedure to fail with the error message "Some level-2 weights vary within group." These 32 schools were therefore excluded to satisfy the assumption of constant higher-level weights required by the WeMix algorithm. After exclusion, the final analytic sample for 2016 consisted of 239 schools, and the two-level random intercept model (students nested within schools) converged without issues. This exclusion was specific to the 2016 cycle and did not occur in the 2011 or 2021 cycles. All models in the three cycles were otherwise specified identically to ensure comparability.

Table 5 Multilevel modelling of reading literacy

	PIRLS 2011		PIRLS 2016		PIRLS 2021	
	Level-1 model	Full model	Level-1 model	Full model	Level-1 model	Full model
	B(SE)	B(SE)	B(SE)	B(SE)	B(SE)	B(SE)
Intercept	277.35*** (9.75)	214.86*** (25.63)	204.20*** (19.87)	160.80*** (39.87)	237.39*** (12.90)	164.01*** (29.99)
Students Like Reading	5.94*** (0.68)	5.94*** (0.69)	4.23* (1.28)	4.18* (1.27)	1.24 (0.88)	1.23 (0.88)
Students Confident in Reading	10.83*** (0.51)	10.81*** (0.51)	16.68**** (0.91)	16.67*** (0.91)	15.46*** (0.68)	15.44*** (0.68)
School Emphasis on Academic Success	-	6.11* (2.48)	-	4.41* (3.13)	-	7.05* (2.28)
ICC	0.397	0.381	0.377	0.375	0.396	0.381

***p < .001, **p < .01, *p < .05

Results of the two-level models are presented in Table 5. Across all three cycles, Students Confident in Reading emerged as the strongest and most consistent predictor of reading literacy, with unstandardized coefficients ranging from 10.81 to 16.68 points (approximately 0.3–0.5 of the PIRLS international standard deviation of 100 points). This effect was highly significant ($p < .001$) in every model and cycle. Students Like Reading showed a positive and significant association with reading achievement in 2011 and 2016 ($B = 5.94$ and $B = 4.23–4.18$, respectively, $p < .05$ to $p < .001$), but the relationship became non-significant in 2021 ($B \approx 1.23$, $p > .05$). School Emphasis on Academic Success exerted a significant positive effect at the school level in all three cycles ($B = 6.11$ in 2011, $B = 4.41$ in 2016, and $B = 7.05$ in 2021; $p < .05$ to $p < .01$), although its magnitude remained relatively modest and somewhat variable across cycles compared to the student-level predictors. All models were specified identically across cycles (same predictors, random intercept only, same weighting approach) to ensure comparability.

The ICCs in Table 5 show substantial between-school variation in reading literacy (0.377–0.397 in Level-1 models; 0.375–0.381 in full models), indicating that 37.5–39.7% of variance was attributable to schools. Compared to null models (Table 4: 38.97–40.60%), the modest reduction in ICCs suggests that student-level attitudes and school emphasis on academic success explained only a small portion of between-school differences, highlighting persistent systemic inequalities in Iran and justifying multilevel modelling. Taken together, the findings reveal a clear temporal shift: whereas students' liking of reading and school emphasis on academic success showed modest but significant effects in earlier cycles, students' confidence in reading consistently emerged as the

single most powerful and stable predictor of reading literacy in Iran across all three PIRLS cycles (2011–2021).

Pairwise comparisons of unstandardised coefficients across consecutive PIRLS cycles showed that the effect of Students Like Reading decreased from 5.94 (SE = 0.69, $p < .001$) in 2011 to 4.18 (SE = 1.27, $p < .05$) in 2016 ($z = -1.22$, $p = .223$), and further to 1.23 (SE = 0.88, $p > .05$) in 2021 ($z = -1.91$, $p = .056$); although the pairwise differences were not statistically significant at the conventional 0.05 level, they indicate a consistent downward trend over time. In contrast, Students Confident in Reading increased significantly from 10.81 (SE = 0.51) in 2011 to 16.67 (SE = 0.91) in 2016 ($z = 5.62$, $p < .001$), then decreased slightly but non-significantly to 15.44 (SE = 0.68) in 2021 ($z = -1.08$, $p = .279$), remaining the strongest and most stable predictor throughout the decade. Meanwhile, School Emphasis on Academic Success exhibited no significant change across cycles—from 6.11 (SE = 2.48) in 2011 to 4.41 (SE = 3.13) in 2016 ($z = -0.43$, $p = .670$) and then to 7.05 (SE = 2.28) in 2021 ($z = 0.68$, $p = .495$)—staying modest and relatively stable.

5. Conclusion

In today's information-driven world, the ability to read and understand written material is a fundamental skill for active societal participation. Developing strong reading skills early is crucial, as it not only fosters a child's cognitive and intellectual growth but also paves the way for success in school, future careers, and life in general (Marôco, 2021). Notwithstanding the critical role of literacy in contemporary societies, the performance of Iranian students in the PIRLS assessments indicates a significant deficiency. Consequently, a more rigorous analysis of the secondary findings from these studies can substantially aid in enhancing students' reading proficiency.

The 29-point decline from 2011 to 2016 corresponds to an effect size (Hedges' g) of approximately 0.29, which is small. The 15-point decline from 2016 to 2021 corresponds to $g \approx 0.15$ (small), and the cumulative 44-point decline from 2011 to 2021 corresponds to $g \approx 0.44$ (approaching medium). These effect sizes indicate that while the declines are statistically significant, their practical magnitude is small to medium. Nevertheless, in educational terms, even a small effect size of 0.15 represents approximately 2–3 months of lost learning (Hill et al., 2008), which is non-negligible when considering the cumulative impact on a generation of students.

Across all three cycles, confidence in reading was the strongest predictor. This effect was more pronounced in 2016 than in 2011 and 2021 (see Table 5). This finding aligns with previous research (Xiao, 2023; Lopes et al., 2022, and others) and the Component Model of Reading (Aaron et al., 2008), which highlights the role of psychological factors. Furthermore, PIRLS 2021 results

showed that students who were "very confident in reading" had an average score of 541, those "somewhat confident" scored 498, and those "not confident" scored 449 (Mullis et al., 2023).

After students confident in reading, school emphasis on academic success had a significant role in predicting achievement in reading in all three cycles. It should be added that comparison of the regression coefficient (β) for the school emphasis on academic success between the 2011 and 2016 cycles shows a decreasing trend. However, this trend reversed in the 2021 study, showing an increase again. Nevertheless, the greatest impact was observed in the 2011 study. This finding was consistent with the results of several research (Nilsen & Gustafsson, 2014; Pongsophon, 2024; Thomure et al., 2025; Costa & Araújo, 2017). In addition, based on PIRLS 2021 results, students in the “very high emphasis” group achieved the highest average score (525), with those in the “high emphasis” group scoring slightly lower (509). The “medium emphasis” group had the lowest average reading achievement, with a score of 486.

Another finding was that "students like reading" was a weaker predictor than reading confidence and school emphasis across all cycles. In PIRLS 2021, it had no meaningful impact on reading achievement, consistent with Taha Thomure et al. (2025) but inconsistent with Lopes et al. (2022). This may be explained by developmental characteristics, the task-oriented nature of school reading, and the rise of smartphones and short-form content (e.g., Reels) in Iran between 2011 and 2021. Thus, enjoyment of informal online reading may no longer align with the comprehension of extended formal texts assessed by PIRLS, weakening the observed relationship.

The other findings indicate that although the average levels of Students Like Reading and Schools' Emphasis on Academic Success remained relatively stable across the examined cycles (see Table 2), their associations with reading literacy showed different patterns over time (see Table 3). The correlation for Students Like Reading declined gradually from 2011 to 2021, whereas the correlation for School Emphasis on Academic Success decreased from 2011 to 2016 and increased again in 2021. This pattern suggests that, in recent years, neither student like in reading nor school emphasis on achievement alone may be sufficient to explain differences in reading performance. Additionally, the evolving nature of students' reading habits and educational priorities may have contributed to the weakening associations. For example, one possible explanation is that students' reading interests may have shifted toward non-academic or digital forms of reading, while schools' priorities may have become increasingly focused on standardized testing rather than fostering deep comprehension and critical literacy. Overall, these findings suggest that internal cognitive and motivational constructs, particularly reading confidence, may play an important role in predicting reading literacy compared with more general factors such as enjoyment of reading or institutional

emphasis on academic success. Another notable finding of this study is the persistently high between-school variance (ICC ≈ 0.38–0.40), indicating that approximately 38–40% of the total variance in reading literacy is attributable to differences between schools, rather than differences among students within the same school. This finding highlights the existence of educational inequality in Iran: schools are not merely reflecting student composition; they are actively contributing to variation in reading outcomes.

PIRLS data show a rising percentage of Iranian students with low reading performance: 25% scored below 400 in 2011, increasing to 35% in 2016 and 40% in 2021. Many fourth-graders struggle with comprehension, inference, and interpretation. However, reading confidence emerged as a decisive predictor of literacy. Teachers can foster confidence by selecting interest-aligned materials, providing gradual success opportunities, and offering encouraging feedback. Educational policymakers can support teachers in this endeavor by providing professional development programs focused on fostering reading confidence and allocating resources for diverse, level-appropriate classroom libraries that enable gradual success experiences. Additionally, school emphasis on academic success matters; education systems should promote high expectations from teachers and parents, convey that everyone can learn, and integrate social skills (e.g., respect, group reading) into school culture. **In** this regard, the education system must invest in teacher training through continuous professional development programs for effective curriculum implementation, create a platform for meaningful parental involvement in students' learning process, and ultimately strengthen school culture with an emphasis on valuing success. It is worth mentioning that a limitation specific to the 2016 cycle was the exclusion of 32 schools (11.8% of the initial school sample) because of non-constant teacher weights within schools. This exclusion was required by the weighting structure of the multilevel analysis in EdSurvey. Although this procedure was necessary for obtaining valid model estimates under the available weighting structure, the possibility of selection bias cannot be ruled out; therefore, the 2016 findings should be interpreted with some caution.

References

Almaskut, A., LaRoche, S., & Foy, P. (2023). *Sample design in PIRLS 2021*. In M. von Davier, I. V. S. Mullis, & B. Fishbein (Eds.), *Methods and procedures: PIRLS 2021 technical report* (pp. 3.1–3.32). TIMSS & PIRLS International Study Center, Boston College. https://pirls2021.org/methods/chap

Bailey, P. & R Core Team. (2023). *EdSurvey: Analysis of NCES Education Survey and Assessment Data* (Version 4.0-1) [R package]. Retrieved from https://www.air.org/project/nces-data-r-project-edsurvey

Bezirhan, U., Foy, P., & von Davier, M. (2023). PIRLS 2021 achievement scaling methodology: Item response theory, population models, and linking across modes. In M. von Davier, I. V.

Cheung, W. M., Lam, J. W., Au, D. W., So, W. W., Huang, Y., & Tsang, H. W. (2017). Explaining student and home variance of Chinese reading achievement of the PIRLS 2011 Hong Kong. *Psychology in the Schools*, 54(9), 889-904.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.

Fishbein, B., Yin, L., & Foy, P. (2023). *PIRLS 2021 User Guide for the International Database*. International Association for the Evaluation of Educational Achievement. https://pirls2021.org/data/downloads/P21_UG_International-Database.pdf

Foy, P., & LaRoche, S. (2020). *PIRLS 2016 user guide for the international database*. TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/pirls2016/international-database/

Fuchs, D., Compton, D. L., Fuchs, L. S., Bryant, V. J., Hamlett, C. L., & Lambert, W. (2012). First-grade cognitive abilities as long-term predictors of reading comprehension and disability status. *Journal of Learning Disabilities, 45*(3), 217–231. https://doi.org/10.1177/0022219412442154

Goddard, R. D., Sweetland, S. R., & Hoy, W. K. (2000). Academic emphasis of urban elementary schools and student achievement in reading and mathematics: A multilevel analysis. *Educational Administration Quarterly*, *36*(5), 683–702

Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 403–422). Mahwah, NJ: Lawrence Erlbaum Associate.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2(3), 172–177.

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge.

Karakolidis, A., Duggan, A., Shiel, G., & Kiniry, J. (2021). Examining educational inequalities: Insights in the context of improved mathematics performance on national and international assessments at primary level in Ireland. *Large-scale Assessments in Education*, 9(1), Article 8. https://doi.org/10.1186/s40536-021-00101-9

Kiamanesh, A., Mohsenpour, M., Safarkhani, M., & Aghdasi, S. (2012). Trends in mathematics achievement of 9th-grade students in Iran from 1999 to 2007 based on findings from the TIMSS international studies and in comparison, with countries in the region, with regard to the goals of the 20-Year Vision Document. *Curriculum Studies*, 6(24), 59–82. [in Persian].

Leino, K., Nissinen, K. & Sirén, M. (2022). Associations between teacher quality, instructional quality and student reading outcomes in Nordic PIRLS 2016 data. Large-scale Assess Educ 10, 25. https://doi.org/10.1186/s40536-022-00146-4

Lopes, J., Oliveira, C., & Costa, P. (2022). School and student determinants of reading performance: A multilevel analysis with Portuguese students. *Revista de Psicodidáctica (English ed.)*, *27*(1), 29-37. https://doi.org/10.1016/j.psicoe.2021.05.001

Marôco, J. (2021). What makes a good reader? Worldwide insights from PIRLS 2016. *Reading and Writing: An Interdisciplinary Journal, 34*(1), 231–272. https://doi.org/10.1007/s11145-020-10068-8

Martin, M. O., Mullis, I. V. S., Arora, A., & Preuschoff, C. (2013). Creating and interpreting the TIMSS and PIRLS 2011 context questionnaire scales. In M. O. Martin & I. V. S. Mullis (Eds.), *Handbook of international large-scale assessment* (pp. 299–318). CRC Press/Taylor & Francis. https://doi.org/10.1201/b16061-19

Martin, M. O., Mullis, I. V. S., Foy, P., & Arora, A. (2012). Creating and interpreting the TIMSS and PIRLS 2011 context questionnaire scales. In M. O. Martin & I. V. S. Mullis (Eds.), *Methods and procedures in TIMSS and PIRLS 2011* (pp. 1–11). TIMSS & PIRLS International Study Center, Boston College. Retrieved from https://timssandpirls.bc.edu/methods/t-context-q-scales.html

Martin, M. O., Mullis, I. V. S., Hooper, M., Yin, L., Foy, P., Fishbein, B., & Liu, J. (2017). *Creating and interpreting the PIRLS 2016 context questionnaire scales*. In M. O. Martin, I. V. S. Mullis, M. Hooper, L. Yin, P. Foy, B. Fishbein, & J. Liu (Eds.), *Methods and procedures in PIRLS 2016* (Chapter 14). TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/publications/pirls/2016-methods/chapter-14.html

Mohammadnia, Z. and Mohsenpour, M. (2025). The Contribution of Student's Variables and Teacher's Variables in Predicting Reading Literacy of PIRLS 2021. *Journal of Educational Sciences*, *32*(2), 29-52. doi: 10.22055/edus.2025.48385.3687 [in Persian].

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). PIRLS 2016 international results in reading. TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/publications/pirls/2016-international-results.html

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (Eds.). (2020). *PIRLS 2021 assessment frameworks*. International Association for the Evaluation of Educational Achievement. https://pirls2021.org/wp-content/uploads/sites/2/2019/04/P21_Frameworks.pdf

Mullis, I. V., & Martin, M. O. (2019). *PIRLS 2021 Assessment Frameworks. International Association for the Evaluation of Educational Achievement*. Herengracht 487, Amsterdam, 1017 BT, The Netherlands.

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

Mullis, S, B. Fishbein, & P. Foy (Eds.), *Methods and procedures: PIRLS 2021 technical report* (pp. 10.1–10.19). Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.tr2103.kb3131

Nilsen, T., & Gustafsson, J. E. (2014). School emphasis on academic success: exploring changes in science performance in Norway between 2007 and 2011 employing two-level SEM. *Educational Research and Evaluation*, 20(4), 308–327. https://doi.org/10.1080/13803611.2014.941371

Nilsen, T., & Gustafsson, J. E. (2016). *Teacher quality, instructional quality and student outcomes: Relationships across countries cohorts and time*. Springer. https:// doi. org/ 10. 1007/ 978-3-319- 41252-8

Ozola, A., & Geske, A. (2019). What do Teachers do to Promote Student' Reading Literacy at 4TH Grade?–Evidence from IEA PIRLS 2016 Study. *Innovations, Technologies and Research in Education*, 323. https://doi.org/10.22364/atee.2019.itre.22

Park, Y. (2011). How motivational constructs interact to predict elementary students' reading performance: Examples from attitudes and self-concept in reading. *Learning and Individual Differences*, 21(4), 347-358. https://doi.org/10.1016/j.lindif.2011.02.009

Park, Y., Chaparro, E., Peciado, J., & Cummings, K. (2015). Is earlier better? Mastery of reading fluency in early schooling. *Early. Educ. Dev., 26*(8), 1187–1209. https://doi. org/10.1080/10409289.2015.1015855

Pongsophon, P. (2024). Deciphering the layers of reading achievement: A multi-level analysis of student and school-level predictors in top-performing PIRLS 2021 countries. *British Educational Research Journal*, 50(6), 2777-2798.

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reza'ee M S, Kiamanesh, Ph.D. A, Naghsh, Ph.D. Z. (2015). The Role of Student and School Variables in Reading Performance among Those Participated in PIRLS 2006. *Education*; 30 (4) :123-140 URL: http://qjoe.ir/article-1-189-fa.html[in Persian].

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). SAGE Publications.

Stiff, J., Lenkeit, J., Hopfenbeck, T. N., and Kayton, H. L. (2023). Research Engagement in the Progress in International Reading Literacy Study: A Systematic Review. (2023). *Educational Research Review*, *40*(5–6), DOI:10.1016/j.edurev.2023.100547

Taha Thomure, H., 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*, 4, 100301.

Tse, S.K., Xiao, Xy. (2014). Differential influences of affective factors and contextual factors on high-proficiency readers and low-proficiency readers: a multilevel analysis of PIRLS data from Hong Kong. *Large-scale Assess Educ* 2, 6 https://doi.org/10.1186/s40536-014-0006-3

Vlachou, A., & Stefanou, C. (2017). School climate and student achievement in PIRLS countries: A multilevel analysis*. Journal of Educational Psychology*, 109(3), 374-389.

von Davier, M., Mullis, I. V. S., Fishbein, B., & Foy, P. (Eds.). (2023). *Creating and interpreting the PIRLS 2021 context questionnaire scales*. In *Methods and procedures: PIRLS 2021 technical report* (Chapter 15). TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.tr2103.kb6994

Wagner, D. (2017). Children's reading in low-income countries. *Read. Teach., 71*(2), 127–133.

Xiao, X.-Y. (2023). How motivational constructs predict reading amount and reading achievement: The role of reading attitude and reading self-concept. *Power and Education*, *15*(3), 341-361. https://doi.org/10.1177/17577438221146246

Yang, G., Badri, M., Al Rashedi, A. and Almazroui, K..(2018). The role of reading achievement: a preliminary examination of fourth graders in Abu Dhabi. *Large-scale Assess Educ* 6, 10. https://doi.org/10.1186/s40536-018-0063-0

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). Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.tr2103.kb6994

Wry, E., & Mullis, I. V. S. (2023). Developing the PIRLS 2021 achievement instruments. In I. V. S. Mullis, M. O. Martin, & B. Fishbein (Eds.), *PIRLS 2021: Technical Report* (pp. 1.1-1.24). Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.tr2101.kb7549

Zare, H. & Naghsh, S. (2016).The Study of Student and Teacher Level on Students' Reading Achievement in Fourth Grade Based on PIRLS Data (2006), *New Approaches in Educational Administration*, 7(26), 39-52. DOR: 20.1001.1.20086369.1395.7.26.3.4[in Persian].