

Do Assessment Practices Really Drive Literacy? A Cross-National Study Reveals Surprising Priorities

Ebrahim Mohammadpour

ebrahim47@gmail.com

Farhangian University

Seifolla Rahmani

Farhangian University

Article

Keywords: student evaluation, evaluation utilization, evaluators, evaluation methods, students assessment practices

Posted Date: April 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-7576025/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Do Assessment Practices Really Drive Literacy? A Cross-National Study Reveals Surprising Priorities

Abstract

This study directly investigates the pivotal question of whether teacher assessment practices are the primary drivers of student literacy. Utilizing 2021 Progress in International Reading Literacy Study (PIRLS) data from Iran, Finland, and Singapore, this cross-national analysis of fourth-grade students reveals a surprising finding: individual student factors (gender, socioeconomic status, and early home literacy activities) exert a far more pronounced influence on literacy scores than the assessment methods themselves. Despite examining a range of common practices—including observations, oral questioning, tests, and projects—the analysis demonstrates that these strategies largely lacked a consistent correlation with literacy performance, challenging conventional assumptions about their effectiveness. A notable exception was project-based assessment in Singapore, which demonstrated a significant positive relationship. The study ultimately reveals a surprising set of priorities: variables such as school emphasis on academic success and teacher gender were identified as more significant contributors to achievement than assessment techniques across the three diverse contexts. These findings advocate for a nuanced understanding of educational effectiveness that prioritizes student-level support and cultural context over a singular focus on assessment tool selection.

Keywords: student evaluation; evaluation utilization; evaluators; evaluation methods; students assessment practices

1. Introduction

Education is pivotal for facilitating learning, a process critically assessed through various evaluation methods to measure student achievement against educational objectives (Saif, 2023). Educational assessment is categorized into three distinct approaches: "of," "for," and "as" learning. Assessment of learning is summative, administered at the end of a program to evaluate progress and eligibility for advancement. In contrast, assessment for learning is an ongoing, formative process that guides instruction and provides feedback to enhance student metacognitive skills (Nikolic et al., 2018; Shao et al., 2024). Assessment as learning empowers students to take ownership of their educational journey, critically self-evaluating and devising strategies for growth (Earl, 2003; Authors, 2018).

Formative assessment, encompassing both "for" and "as" learning, provides crucial data for tailoring instruction to individual student needs (Chen et al., 2021), facilitating adjustments that enhance academic performance (Dunn & Mulvenon, 2009). This study directly interrogates the assumed primary role of assessment practices in driving fourth-grade reading literacy. It explores how demographic factors (gender, socioeconomic status-SES) and teacher characteristics (experience, satisfaction) interact to influence literacy outcomes across different cultural contexts. The paper seeks to answer: how much do assessment practices really drive literacy, and what factors truly matter?

2. Significance and Rational of the Study

This research critically examines the presumed powerful link between teacher assessment practices and reading literacy, testing it against the influence of student-level and systemic factors. It employs a "contrast of contexts" approach (Bray & Thomas, 2007), selecting Iran, Finland, and Singapore as archetypal cases representing vastly different educational philosophies, cultural priorities, and performance levels. This design moves beyond single-case studies to compare assessment practices in two high-achieving but distinct systems with a lower-performing nation facing structural challenges. The aim is to test the universal efficacy of common assessment practices and identify contextually effective factors for enhancing reading literacy.

2.1 Theoretical Grounding: The "Contrast of Contexts" Framework

The selection of Iran, Finland, and Singapore is framed within comparative education's "contrast of contexts" approach. Stark contrasts amplify the influence of systemic factors,

revealing how policy, culture, and institutional capabilities interact to influence outcomes. This allows the research to ask if a practice effective in a high-performing system yields the same result in a system with different challenges. Iran's challenges in international assessments (Mullis et al., 2023), marked by centralized governance and teacher-centered pedagogies (Sadeghi et al., 2020), stand in stark contrast to Finland's decentralized, trust-based framework and Singapore's innovative hybrid model. Juxtaposing these systems contributes to a relational understanding of educational ecosystems.

2.2 Empirical and Pragmatic Justification for Country Selection

Empirically, the three countries present a clear gradient in PIRLS performance, the study's dependent variable. They differ maximally on key independent variables: governance (Iran: centralized; Finland: decentralized; Singapore: centralized with flexibility), pedagogical focus (Iran: teacher-centered; Finland: student-centered; Singapore: hybrid), and socio-cultural support. Pragmatically, the availability of comparable PIRLS 2021 data for all three ensures methodological rigor.

2.3 Iran as a Critical Case: Addressing a Gap in Comparative Research

Including Iran addresses a gap in comparative literature, which often overlooks low-performing contexts. This prevents a selection bias where "best practices" are derived only from successful cases. Iran's inclusion tests prevailing notions and highlights the necessity for policies sensitive to local inequities, such as insufficient training in formative assessment (Carnoy, 2019).

3. Overview of Assessment Practices (Contextualized within the Chosen Frameworks)

Assessment practices are not universal; their implementation is shaped by broader educational context. This study focuses on five key assessment practices from the PIRLS 2021 framework: observing students, asking oral questions, short regular written assessments, longer tests/exams, and long-term projects. These represent a spectrum from formative to summative purposes.

3.1 Defining the Five Core Practices:

- *Observational Assessment:* This fundamental approach provides teachers with immediate insights into student engagement, comprehension, and skill application. Structured observation enables tailored instruction and targeted interventions, aligning with the

formative assessment principles of continuous feedback (Black & Wiliam, 1998; Hunter & Bailey, 2020).

- *Questioning Techniques*: The practice of encouraging student responses is essential for formative assessment. Well-structured questioning enhances cognitive processing, offers immediate insights into understanding, and promotes critical thinking by creating a dynamic classroom environment (Black & Wiliam, 1998; Hernando et al., 2023).
- *Regular Written Assessments*: These frequent, short-cycle assessments—whether on paper or digital—provide consistent insights into student progress. They are a cornerstone of a holistic assessment framework, guiding instructional methods and fostering student self-assessment (Popham, 2008). Digital tools further enhance this by streamlining data collection and analysis (Eden et al., 2024).
- *Longer Tests and Examinations*: fulfilling a summative role, these assessments facilitate a comprehensive evaluation of knowledge retention across broader content. Their effectiveness is contingent on careful design that aligns with learning objectives to ensure they reliably measure intended outcomes (Briggs, 2010; Gibson & Graham, 2016).
- *Projects*: Long-term projects encourage deep engagement with material and enhance critical thinking. By requiring ongoing work like reading logs, they promote personal reflection, autonomy, and accountability, connecting theoretical learning to practical application and nurturing intrinsic motivation (Deci & Ryan, 2000; Gordon, 2019; Wiggins & McTighe, 2005).

3.2 Identifying the Research Gap:

The introduction of these strategies in PIRLS 2021 reveals a significant gap: a deficiency in comparative studies analyzing their effectiveness between high- and low-performing nations. This study addresses this gap by systematically examining whether the relationship between these strategies and reading literacy is universal or context-dependent.

Research Questions:

1. How do these assessment strategies influence fourth-grade reading literacy across diverse educational contexts?

2. How do student factors, teacher attributes, and school environments shape reading literacy outcomes?
3. How do specific teacher assessment strategies, accounting for teacher backgrounds, affect reading literacy performance across nations?
4. What insights emerge from comparing how different countries' assessment strategies contribute to reading literacy?

4. Conceptual Framework

Assessment practices are conceptually associated with multiple factors at student, classroom, and school levels (Chen et al., 2021).

4.1 Student-Level Factors

Individual differences significantly influence reading achievement. Girls typically outperform boys, a disparity linked to sociocultural influences and attitudes (Voyer & Voyer, 2014). Students from lower SES backgrounds face obstacles to literacy growth; the correlation between SES and performance underscores an educational gap (Bradley & Corwyn, 2002; Willms, 2018). According to family investment theory (Conger & Donnellan, 2007), higher-SES parents provide more resources for development (Bergen et al., 2017). Early literacy activities at home and positive student attitudes toward reading are also critical for proficiency (Mullis et al., 2007; Toste et al., 2020).

4.2 Classroom-Level Factors

Teacher experience correlates with effective instructional and assessment strategies (Darling-Hammond, 2000). Teacher job satisfaction is linked to more dynamic and effective assessment practices that foster a growth mindset (Caprara et al., 2006). Constructive feedback during formative assessments enhances self-efficacy, motivation, and academic advancement (Zumbrunn et al., 2016). Teacher practices contribute significantly to variance in reading scores (Torres et al., 2021).

4.3 School-Level Factors

School climate significantly influences academic achievement (Scherer & Nilsen, 2016). A school's academic focus, characterized by high expectations and data-driven instruction, is closely linked to student motivation and success (Hattie, 2009). When educational leaders demonstrate commitment to excellence, it fosters a collective vision, reinforcing practices like differentiated

instruction (Darling-Hammond et al., 2017). In high-achieving systems like Singapore, coherence between national policies and school practices facilitates improvement (Deng et al., 2020).

5. The Educational Backgrounds of the Studied Countries

Iran's centralized system, overseen by the Ministry of Education, introduced descriptive evaluation in 2002 to improve learning outcomes and critical thinking (Authors, 2018).

Finland considers education a fundamental right, ensuring equal opportunities. Its decentralized system provides nine years of free basic education with a focus on holistic development (Finnish National Agency of Education, 2017).

Singapore's system is globally recognized for its high performance. It has shifted from a focus on performance to a more holistic approach encouraging critical thinking and national-level learning (Ministry of Education Singapore, 2023). Unlike Iran and Finland, Singapore has historically grouped students by ability.

6. Methodology

6.1 Data

The 2021 PIRLS data used in this study was obtained from the TIMSS and PIRLS Center (<https://timssandpirls.bc.edu/>). This study focused on fourth grade students from Iran, Finland, and Singapore, with a total of 5962, 7018, and 6719 students respectively. These students were from 218, 219, and 183 schools in the respective countries. The average age of students in Iran, Finland, and Singapore is 10.2, 10.8, and 10.4 years respectively.

6.2 Measures

Reading achievement: The PIRLS Center collected data on student performance in the PIRLS 2021 literacy assessment, which evaluates reading and comprehension in informational and literary domains. Students' responses impact their overall scores and provide insights into specific areas, leading to additional scores for literary, informational, interpretive, and direct processes.

The PIRLS studies employed a matrix sampling or rotated-booklet design, allowing each student to respond to a certain number of items from a large number of reading items. In contrast, PIRLS aimed to provide individual students with a comprehensive score for the entire test, leading to the generation of five '*plausible values*' representing a student's potential performance across all items. These plausible values serve as estimates of how students might have scored had they completed the full set of items. PIRLS utilizes this plausible values methodology alongside other

scoring metrics (Hastedt, 2006). Consequently, the plausible values for each scale represent the most accurate indicators of student achievement within the PIRLS dataset and should be utilized as the primary outcome variable in any research concerning student performance (Foy & Olson, 2009). To estimate overall reading achievement, this study utilized the average of five plausible values—a standard approach in large-scale assessments to account for measurement error (OECD, 2019).

Socioeconomic Status (SES): Students' SES was evaluated using four ordinal indicators derived from the PIRLS parent questionnaire: (1) number of books in the home, (2) number of children's books, (3) highest parental education level, and (4) highest parental occupation level. Following the PIRLS International Study Center's standardized methodology (Mullis et al., 2023, p. 88), these indicators were aggregated into a composite SES variable and categorized into three discrete groups (low, medium, high) to align with cross-national benchmarks and avoid misrepresenting ordinal data as continuous. The categorization criteria are as follows:

- *High SES* (cut score ≥ 11.1): Households with over 25 books and children's books, at least one parent holding a university degree, and at least one parent employed in a professional occupation.
- *Low SES* (cut score < 8.5): Households with 25 or fewer books and children's books, parents without education beyond upper secondary level, and occupations classified as unskilled or skilled workers.
- *Middle SES*: All students not meeting the criteria for high or low SES are classified into this middle category, ensuring a comprehensive representation of socioeconomic diversity.

This tripartite classification ensures methodological consistency with PIRLS reporting practices while providing actionable insights into how specific dimensions of SES (e.g., cultural capital, resource access) correlate with literacy outcomes.

Home early literacy activities before primary school: This factor was assessed through parental perceptions of nine specific reading activities. The PIRLS Center classified children into three distinct groups based on their engagement levels. Often, refers to those who regularly participated in early literacy activities tended to achieve scores at or above the established cut score, which aligns with parental feedback indicating frequent involvement in five of the nine activities and occasional participation in the remaining four. This group had a cut score of 10.7 or higher. Never or almost never, students who rarely or never engaged in these activities scored at

or below the cut score, reflecting parental reports of infrequent involvement in five activities and occasional engagement in the others, with scores ranging from 6.2 to 10.7. A third group, sometimes, consisted of students who sometimes engaged in early literacy activities, characterized by a cut score below 6.2.

Student gender: The student questionnaire identified gender, with 1 for girls and 2 for boys.

Assessment practices: The PIRLS framework proposed five assessment strategies of (1) observing students as they work, (2) asking oral questions, (3) conducting brief tests, (4) performing long tests, and (5) assigning long-term projects. Teachers rated these strategies on a three-point scale, where 1 indicated 'very much', 2 signified 'moderate', and 3 represented 'not at all'.

The study examined assessment strategies by analyzing the relationships among factors such as teachers' gender, age, experience, qualifications, job satisfaction, and students' reading performance. Data were from the PIRLS teacher questionnaire, with '*gender*' coded as 1 for female and 2 for male. '*Age*' was categorized as: 1 (under 25), 2 (25-29), 3 (30-39), 4 (40-49), 5 (50-59), and 6 (60+). '*Educational attainment levels*' included: 1 (incomplete high school), 2 (high school), 3 (pre-university), 4 (short university course), 5 (bachelor's), 6 (master's), and 7 (doctorate).

Teacher job satisfaction: The PIRLS Center assesses this factor using a five-indicator scale that includes satisfaction with the profession, work meaningfulness, passion, motivation, and pride. Teachers' feedback categorizes students into three groups based on a four-point Likert scale from "always" to "never or almost never." Students are classified as "high satisfaction" if their teachers rated "always" for three indicators and "often" for two. Those in the "dissatisfied" group had teachers who selected "sometimes" for three indicators and "often" for the other two, while the remaining students were taught by teachers with "somewhat" satisfaction.

Principals' perceptions of students' academic success: The PIRLS Center classifies schools into three categories of "very high emphasis," "high emphasis," and "moderate emphasis." This classification evaluates twelve dimensions of the school's commitment to academic achievement. In the "very high emphasis" category, students rate six dimensions as highest and the other six as high. In the "moderate emphasis" category, they rate six dimensions as medium and the other six as high. The remaining schools fall under "high emphasis." Principals used a five-point Likert scale, where 1 indicates very high emphasis and 5 indicates very low emphasis.

6.3 Analysis Technique

The study utilized multi-level modeling, with student factors at level 1 and a combination of class and school factors at level 2. PIRLS employed a multi-stage cluster sampling design, resulting in unequal probabilities of selecting schools and students within each country. To prevent bias in parameter estimation, weighting variables were developed by the PIRLS Center, including "total student weight" for student level and "teacher weight" for school level. In cross-country research where sample sizes vary significantly, employing total student weight can lead to misleading interpretations, as it applies a uniform approach to these disparities. To mitigate this concern, the PIRLS Center developed two specific sampling weights, SENWGT and HOUWGT, designed for cross-national comparisons (Foy & Olson, 2009; Rutkowski et al., 2010). In the current study, "SENWGT" was applied for analyses at the student level, and "school weight factor" was used as the weighting variable for school-level assessments.

The parameters were estimated through the full maximum likelihood method (Hox, 2009), while predictor factors were determined using the grand mean centering method (Enders & Tofghi, 2007). Data preparation was done using SPSS 26 software, and integration was carried out using IDB Analyzer. The primary data analysis was then conducted using HLM 6.07.

6. Results

Five models are estimated for each country as follows:

- 1) Null or Unconditional model: In this model, there is no predictive factor at the student or school level.
 - a. **Level-1 model:** $Y_{ij} = \beta_{0j} + r_{ij}$ $r_{ij} \sim N(0, \sigma^2)$ $\sigma^2 = \text{level-1 variance}$
 - b. **Level-2 model:** $\beta_{0j} = \gamma_{00} + u_{0j}$ $u_{0j} \sim N(0, \tau_{00})$ $\tau_{00} = \text{level-2 variance}$
- 2) Model 1: All student level factors are considered as fixed effects in this model.
- 3) Model 2: Only five assessment strategies, which are part of the school level factors, were included.
- 4) Model 3: This model includes all school level factors, including assessment strategies, together.
- 5) Final (Full) model: This model incorporates all factors from both student and school levels.

7.1 The Null Model

The results indicate that the variability of reading scores among fourth grade students in Iran, Finland, and Singapore can be attributed to differences between schools, with 46.07%, 11.95%, and 20.99% of the variation being explained, respectively. The amount of variance explained by this model can be calculated by subtracting the variance of Model 1 from the Null model and then dividing by the Null model.

The proportion of variance explained by each model is visually summarized in Figure 1, which highlights the incremental explanatory power of predictors across analytical tiers. This visualization underscores how school-level factors, when integrated with student-level variables, enhance the model's capacity to account for cross-national differences in reading achievement.

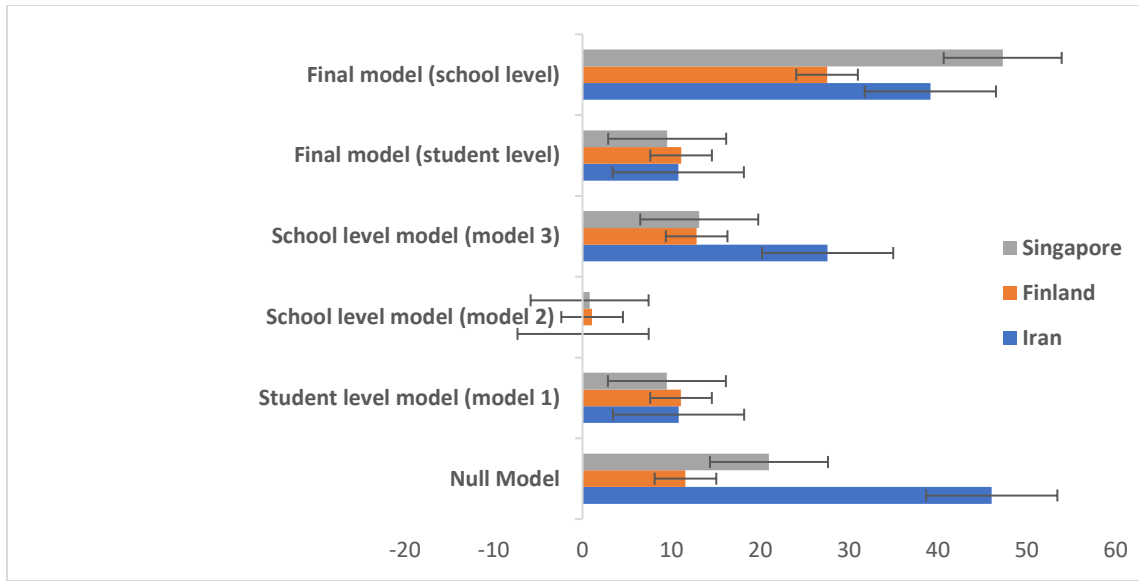


Figure 1: Comparative variance explained by student-, school-, and full-model specifications

The data presented on Figure 1 indicates that factors related to students exerted a nearly equivalent influence on reading scores across the three countries. Nevertheless, student-level factors contributed more significantly to the variations observed at the school level for Iranian and Singaporean students than for their Finnish counterparts. Model 3 provided a more robust explanation for the disparities at the school level among Iranian students in comparison to those from Finland and Singapore.

7.2 Student-Level Predictors of Reading Achievement (Model 1)

Model 1 focused on analyzing the correlation between the student-level factors and students' reading literacy performance and the results are presented in Table 1.

Table 1*The correlation between student-level factors and reading scores*

Factor	Iran	Finland	Singapore
Reading scores	400.55** (7.55)	549.46** (2.04)	582.10** (2.32)
Student gender	-41.44** (8.95)	-17.84** (2.23)	-13.18** (2.07)
Students SES	13.47** (1.11)	12.53** (0.74)	14.82** (0.75)
Home early literacy activities before primary school	4.87** (0.73)	3.79** (0.69)	3.42** (0.39)

(*) $p < .05$ and (**) $p < .001$

The correlation among the three factors and reading literacy is statistically significant across all countries ($p < 0.001$). Girls outperformed boys in all three countries, with the gender gap being more pronounced in Iran compared to Finland and Singapore. The association between SES and reading literacy is comparable for Iranians and Finns, but slightly stronger for Singaporeans. Conversely, Iranian home early literacy activities before primary school show a stronger correlation with their literacy achievement than in the other two countries.

7.3 School-Level Predictors of Reading Achievement

Model 2 (Assessment Strategies Only): The correlation between assessment strategies and students' reading literacy performance was assessed by analyzing the five strategies independently before adjusting for teacher background variables, with the findings presented in Table 2. Notably, the model accounted for a higher percentage of variance among Iranians compared to the other two countries.

Table 2*The correlation between the assessment strategies and students' reading literacy performance*

Factor	Iran	Finland	Singapore
Reading scores	395.32** (8.29)	548.14** (2.02)	581.92** (2.90)
Observing students as they work	-34.33 (32.56)	4.88 (5.17)	9.62 (6.06)
Asking students to answer questions during class	-1.20 (32.56)	-7.09 (4.72)	-11.80 (9.18)
Short, regular written assessments (paper or digital)	-1.81 (29.86)	-3.94 (3.22)	3.22 (6.83)
Longer tests (e.g., unit tests or exams)	-9.14 (13.47)	-4.87 (4.40)	2.41 (5.23)
Long-term projects (e.g., reading logs)	6.89 (15.95)	3.46 (3.10)	10.83* (4.90)

(*) $p < .05$ and (**) $p < .001$

In a key finding that challenges conventional wisdom, the data indicates no significant correlation between the five assessment strategies and students' literacy performance in the three countries, except for the project method, which showed a significant relationship in Singapore ($p < 0.05$).

Model 3 (All School-Level Factors): the relationship between all the school-level factors (including the assessment strategies) was examined and the results are detailed in Table 3.

Table 3

The correlation between the all the school-level factors and students' reading scores

Factor	Iran	Finland	Singapore
Reading scores	406.52** (5.67)	548.18** (1.87)	581.92** (2.68)
Observing students as they work	9.90 (26.59)	-0.01 (0.27)	8.99 (5.80)
Asking students to answer questions during class	-18.30 (24.90)	-2.12 (4.56)	-6.79 (9.57)
Short, regular written assessments (paper or digital)	8.34 (16.96)	0.70 (2.69)	9.08 (6.65)
Longer tests (e.g., unit tests or exams)	-6.15 (12.61)	-2.62 (1.41)	0.76 (5.00)
Long-term projects (e.g., reading logs)	5.14 (13.98)	2.82 (4.35)	8.26* (4.00)
Teachers teaching experience	0.06 (1.20)	-0.72 (4.28)	0.89 (0.55)
Teacher's gender	-63.39** (14.23)	-3.40 (3.27)	-13.35 (9.02)
Teacher's age	-4.43 (10.96)	-5.71 (3.32)	-7.44 (5.01)
Teacher's level of education	-11.89 (8.54)	2.49 (3.03)	3.16 (4.57)
School emphasis on academic success	5.56 (3.23)	5.23** (1.36)	7.67** (1.42)
Teachers job satisfaction	4.67 (4.21)	0.90 (0.88)	-0.95 (1.01)

(*) $p < .05$ and (**) $p < .001$

Results from Table 3 show a significant link between using the project method to evaluate students and literacy outcomes in Singapore, even after adjusting for background variables ($p < 0.5$). Conversely, female teachers in Iran outperform male students significantly ($p < 0.001$). The association between school focus on academic success and reading literacy performance is robust in Singapore and Finland ($p < 0.001$), but there was no significant association between school focus on academic success and reading literacy performance for Iranian students. Factors like

teachers' experience, age, education level, and job satisfaction did not correlate significantly with literacy performance across the countries.

7.4 The Full Model (Model 5)

The Full model analyzed the relationship between student-level and school-level variables and reading literacy achievement. The findings indicated (Table 4) that the influence of student-level variables on reading literacy performance was nearly identical across the three countries after controlling for school-level factors. Conversely, school-level variables had a greater impact on reading literacy performance for Singaporeans and Iranians compared to Finns when adjusting for student-level variables.

Table 4

The correlation between the student- and school-level factors with students' reading scores

Factor	Iran	Finland	Singapore
Reading scores	407.45** (5.20)	549.24** (1.86)	582.11** (2.14)
Student gender	-38.78** (9.13)	-17.82** (2.81)	-13.11** (2.08)
Students SES	13.14** (1.10)	12.48** (0.75)	14.67** (0.86)
Home early literacy activities before primary school	4.92** (0.72)	3.87** (0.70)	3.48** (0.39)
Observing students as they work	1.14 (24.28)	0.01 (0.21)	6.86 (4.76)
Asking students to answer questions during class	-12.69 (23.70)	-0.97 (4.21)	-5.22 (7.93)
Short, regular written assessments (paper or digital)	6.90 (15.94)	0.37 (4.56)	7.08 (5.38)
Longer tests (e.g., unit tests or exams)	-5.87 (11.04)	0.18 (1.30)	1.92 (4.09)
Long-term projects (e.g., reading logs)	6.00 (12.48)	5.10 (4.28)	6.95* (3.40)
Teachers teaching experience	0.16 (1.10)	-1.80 (3.80)	0.54 (0.46)
Teacher's gender	-37.40** (12.78)	-6.37* (2.93)	-9.90 (7.39)
Teacher's age	-4.30 (11.27)	0.10 (3.05)	-5.31 (4.25)
Teacher's level of education	-9.25 (7.91)	3.40 (3.00)	2.54 (3.48)
School emphasis on academic success	3.66 (3.00)	3.80* (1.28)	5.53** (1.12)
Teachers job satisfaction	4.20 (3.81)	0.97 (0.76)	-0.67 (0.84)

(*) $p < .05$ and (**) $p < .001$

After accounting for school-level factors, all three student-level factors were strongly correlated with reading literacy scores in the three countries ($p < 0.001$). The relationship between

the scores of male and female students in Iran was significantly stronger compared to that of their counterparts in Finland and Singapore, where the correlation was nearly equivalent. Student SES and home early literacy activities showed nearly identical correlations with reading literacy across all three countries.

When considering student-level factors, school-level factors like teacher's assessment strategies and backgrounds did not show significant correlations with reading literacy. Notably, however, the project method in Singapore demonstrated a significant relationship ($p < 0.05$), as did the gender of teachers in Iran ($p < 0.001$) and Finland ($p < 0.05$). Furthermore, the emphasis placed on academic achievement by schools was significantly associated with student literacy outcomes in Finland ($p < 0.05$) and Singapore ($p < 0.001$), while this was not the case for students in Iran.

7. Discussion

8.1 Cross-National Patterns in Literacy Determinants

The comparative analysis of literacy achievement in Iran, Finland, and Singapore yields significant insights into the determinants of educational outcomes. Notably, in Iran, student-level factors—such as motivation, ability, and family background—have a more profound impact on literacy than school-level factors. This aligns with previous research indicating a strong correlation between individual characteristics and academic performance (Steinmayr et al., 2019).

A key finding is the considerable variability in literacy scores among Iranian students, contrasting with the more consistent results in Finland and Singapore. This variability is linked to school-level factors like administrative efficiency, school climate, resource availability, and diverse student demographics. Research consistently shows that students in well-organized educational environments achieve superior academic outcomes compared to those in less supportive conditions (McNaughton et al., 2013; Hattie, 2009; Rimm-Kaufman & Sandilos, 2018). Consequently, Iranian schools exhibit a wide range of conditions and resources, while Finnish and Singaporean educational systems display greater uniformity.

The examination of reading skills indicates that literacy disparities extend beyond school effectiveness, reflecting significant variations in individual and family backgrounds. Contextual theory (Robinson, 1950) emphasizes the complexity of assessing educational performance, shaped by these diverse influences. The marked effect of family background in Iran underscores an educational "learning divide" that influences literacy outcomes.

The analysis highlights considerable differences in literacy levels among the three countries, with Iranian students displaying a wide range of scores, while Finnish students show relatively consistent performance. School quality has a distinct impact in Iran and Singapore, while Finnish schools emphasize educational and social equity. Notably, when considering school-level factors, the explainable variance in literacy scores for Iranian students decreases, whereas it increases for Finnish students and remains stable for Singaporean students. This suggests that school-level factors play a critical role in literacy outcomes, particularly in Singapore, where roughly 50% of the variance in literacy scores is attributable to these factors.

The significant influence of individual student factors, including motivation, capability, and family background, on literacy results in Iran calls for a focused intervention strategy. It is essential for policymakers to emphasize initiatives that boost student motivation and assist families in creating supportive learning atmospheres. Implementing programs that promote parental involvement, offer psychological support to students, and develop customized motivational approaches can play a crucial role in closing the educational gap. Furthermore, in stark contrast, the relatively uniform literacy levels observed in Finland and Singapore exemplify the efficacy of organized and equitable educational systems. For countries like Iran, these models suggest that a shift towards prioritizing educational and social equity is imperative. Policymakers should assess existing school structures, aiming to implement strategies reflective of the Finnish model, which emphasizes both educational quality and social support systems.

8.2 Gender and Socioeconomic Disparities in Reading Achievement

Gender disparities in reading performance present a significant challenge in educational analysis, particularly evident in the comparison of outcomes across different countries. Findings indicate that girls outperform boys in reading, a trend consistent with global academic performance patterns (Mullis et al., 2023). The performance gap is especially pronounced in Iran, where boys lag considerably behind girls, more so than Finland and Singapore. This discrepancy is believed to arise from differing motivational beliefs and attitudes toward reading (Mullis et al., 2007; Kush et al., 2005; Toste et al., 2020). Girls typically demonstrate a greater enthusiasm for reading, which correlates with higher academic engagement (OECD, 2019). Conversely, boys often show less commitment, resulting in reduced study time and inferior academic outcomes.

The complexity of these performance differences reflects the interplay of numerous factors at both individual and school levels. While the school environment is critical, factors such as student

motivation, family background, and gender dynamics significantly shape literacy achievements. Understanding these influences is essential for educators and policymakers who aim to improve literacy outcomes. Addressing ongoing educational inequalities, particularly the striking gender gaps and variations among school types, necessitates targeted strategies to enhance student motivation and optimize resource allocation to achieve more equitable literacy outcomes in Iran.

Further exploration reveals the roots of these disparities. Girls' strong reading skills in early education stem from intrinsic motivation and favorable attitudes towards literacy (OECD, 2019). Research has shown that adolescent attitudes toward reading are paramount in understanding gender differences (Acar-Erdol & Akin-Arikan, 2022; Heyder et al., 2017). Boys' lower reading engagement can be partly attributed to the perception that reading is primarily a female activity, which reinforces negative stereotypes and reduces their motivation (Heyder & Kessels, 2017; Muntoni et al., 2021).

Moreover, school environments significantly impact reading engagement. Studies have indicated that boys' perceptions of educational settings greatly influence their reading skills, while girls thrive in supportive academic atmospheres (Logan & Johnston, 2009). Although schools can mitigate gender gaps in literacy outcomes (Downey et al., 2022), individual attitudes towards academic subjects remain critical to performance (Van Houtte, 2004). Additionally, extracurricular reading habits further perpetuate performance differences, as girls tend to engage more in reading outside the classroom, which fosters a stronger enthusiasm for literacy (Smith & Reimer, 2023).

Addressing gender disparities in reading performance is critical for educators and policymakers, particularly in contexts like Iran, where boys often lag behind girls academically. To mitigate this gap, educational initiatives should focus on identifying the root causes of boys' underperformance and implementing strategies to enhance their reading motivation. This includes challenging societal stereotypes that link reading with femininity by promoting male role models and establishing boys-only reading groups.

Furthermore, schools must cultivate inclusive environments through gender-sensitive teacher training and differentiated instructional methods that engage all students effectively. Involving families is also essential; workshops for parents can underscore the importance of reading and encourage shared reading practices at home. Additionally, expanding extracurricular offerings—such as book clubs or reading challenges specifically designed for boys—can significantly boost

their interest in literacy. Research supports that participation in engaging reading activities fosters intrinsic motivation, leading to improved academic outcomes (Gambrell & Marinak, 2010).

Finally, continuous research and ongoing data collection regarding gender disparities in reading are vital. Policymakers should conduct regular assessments to monitor progress and adapt strategies according to students' evolving needs, ensuring that resources are allocated efficiently and interventions remain effective.

The relationship between SES and reading literacy presents a complex situation that warrants careful consideration. Research conducted by Larson et al. (2015) and further investigations underscore a significant correlation between family SES and students' reading achievements across diverse cultural contexts. Notably, the impact of SES on literacy often outstrips that of early home literacy practices, underscoring the crucial role of family support in the development of reading skills (NAEYC, 2018). For instance, findings indicate that Iranian students display a considerable connection between literacy scores and early reading activities, highlighting the disparities in family involvement linked to varying SES levels (Mullis et al., 2023).

Moreover, the educational environment reveals significant disparities in student performance influenced by teacher gender. In particular, an analysis of student outcomes in Iran, Finland, and Singapore demonstrates that Iranian students perform better when instructed by female teachers. In contrast, the gender of teachers in Finland and Singapore does not significantly affect student achievements. This variation calls for a comprehensive examination of the cultural, systemic, and pedagogical factors that drive these educational outcomes.

The superior performance of Iranian students under female teachers aligns with broader patterns in gender-differentiated teaching practices observed in Middle Eastern contexts (Asadullah & Chaudhury, 2020). This phenomenon may reflect systemic gender norms where female educators often face societal expectations to adopt nurturing, student-centered roles—a pattern documented in Iran's teacher training programs (Mehran, 2022). By contrast, Finland's gender-neutral outcomes align with its institutional commitment to equity, where teacher training explicitly emphasizes pedagogical competence over demographic characteristics (Sahlberg, 2015; Kumpulainen et al., 2020).

While it is important to acknowledge the contributions of female teachers, it is equally vital to recognize the presence of exceptional male teachers (American Educational Research Association, 2018). The variations in student performance across different educational systems

stem not solely from teacher gender but from a complex interplay of cultural norms, teaching methodologies, and institutional priorities.

In Singapore, the educational system demonstrates a strong correlation between a school's emphasis on academic success and student performance; however, this relationship diverges significantly from that observed in Iran and Finland. In Iranian schools, a competitive atmosphere—especially in subjects such as mathematics and science—may lead to an overemphasis on academic results, potentially stifling the development of essential practical skills necessary for future careers (Tan, 2018). In contrast, schools in Singapore employ diverse assessment strategies, including project-based learning, which have proven effective in improving reading literacy outcomes, particularly for Iranian students (Ho & McBride, 2020).

Insights into the relationship between teacher gender and student performance highlight the need for strategic teacher recruitment and training. Evidence from Iran shows that female teachers significantly enhance student outcomes, underscoring the importance of professional development to engage and retain them. Simultaneously, encouraging male teachers to take active roles is essential for creating an equitable learning environment. Continuous training focused on dedication and pedagogical excellence should be provided for all educators, regardless of gender. The varied results in educational systems like Iran, Finland, and Singapore suggest that policies should promote diverse teaching methods and assessment techniques. Iranian practices, which often emphasize competition, could benefit from adopting more holistic approaches seen in Finland and Singapore, such as collaborative learning and project-based assessments, to improve student skills and reading literacy. Finally, educational policy must consider context when addressing socioeconomic disparities and the impact of teacher gender. Policies should focus not only on teacher demographics but also on creating inclusive, culturally responsive teaching environments that meet the diverse needs of all students, ensuring equitable opportunities for literacy and success.

The current research examines the relationship between school-level contextual factors and reading literacy assessment strategies among fourth-grade students in Iran, Singapore, and Finland. The characteristics of teachers, such as age, experience, and educational qualifications, demonstrated a minimal correlation with the employed assessment strategies. Notably, only in Iran was there a significant connection between teacher gender and student literacy outcomes.

The patterns observed may indicate the limitations associated with self-reported data, particularly the influence of social desirability bias, which is a well-recognized issue in cross-cultural educational research (Nederhof, 1985). Teachers' answers regarding assessment strategies might be swayed by institutional or societal norms surrounding "ideal" teaching practices, potentially masking the true connections between contextual factors and literacy outcomes. To improve the validity of future research, it is advisable to complement teacher surveys with classroom observations, a method that has proven effective in large-scale assessments like PISA (OECD, 2019). This combined approach would yield a more accurate depiction of classroom practices and help reduce response bias.

The uniformity of assessment strategies across genders in Finland and Singapore may reflect broader global trends in teacher professionalization (Darling-Hammond et al., 2017). However, Iran's outlier status underscores the persistence of localized gender norms in shaping classroom practices (Asadullah & Chaudhury, 2020).

Singaporean teachers' preference for project-based assessments (PBAs) aligns with their educational system's documented prioritization of critical thinking and academic achievement (Tan, 2018), a framework associated with enhanced literacy outcomes. In contrast, Iranian and Finnish pedagogical traditions—rooted in centralized curricula (Ghorbani & Alavi, 2018) and equity-driven practices (Sahlberg, 2015), respectively—reflect differing sociopolitical values that shape assessment strategies. The limited efficacy of PBAs in Iran resonates with Bronfenbrenner's (1979) ecological systems theory, which posits that educational practices derive meaning from their alignment with societal values. Culturally responsive pedagogy (Banks, 2008) further underscores the necessity of adapting PBAs to community-level priorities, particularly in contexts where rote memorization dominates instructional norms.

The weak correlation between teacher characteristics—like age, experience, and qualifications—and their assessment strategies suggests that current professional development may not meet teachers' diverse needs. Policymakers should create tailored training programs that promote varied assessment techniques, considering each nation's sociocultural context, with a focus on innovative methods to enhance students' reading literacy. Additionally, the potential for response bias in self-reported data raises concerns about the reliability of educational studies. To improve accuracy, educational authorities should foster a culture of openness and self-reflection among teachers, implementing anonymous feedback and peer assessments for better insights into

assessment practices. Furthermore, Singaporean teachers' preference for project-based assessments, linked to improved literacy outcomes, offers a valuable approach that could be adapted in Iranian and Finnish educational contexts. These countries should explore integrating project-based learning while considering their unique socio-political environments to promote critical thinking and creativity—key skills in today's global landscape.

8.3 Project-Based Assessments: Cultural Alignment and Challenges

The significant relationship between project-based assessments (PBAs) and student reading performance in Singapore, as opposed to its limited impact in Iranian and Finnish educational contexts, can be attributed to various factors inherent to each country's educational philosophy, cultural values, and teaching pedagogical approaches.

PBAs are increasingly valued within Singapore's educational framework for their role in cultivating vital 21st-century competencies such as critical thinking, creativity, and collaboration (Ministry of Education, 2015). Unlike traditional assessments, which predominantly emphasize memorization and recall, typically utilizing paper-and-pencil tests or digital formats, PBAs prioritize the practical application of knowledge to solve real-world problems (Tan, 2018). These assessments can manifest in diverse formats, including research initiatives, presentations, and prototypes, with an emphasis on the entire project lifecycle—encompassing planning and teamwork—rather than solely the end result (Koh, 2017). Furthermore, PBAs promote interdisciplinary learning and utilize transparent rubrics to ensure consistent evaluation (Au, 2020).

The advantages of PBAs encompass a deeper comprehension and engagement through active involvement, enhanced critical thinking and collaboration skills, and greater student ownership of their educational journey (Tan, 2018; Ministry of Education, 2015). They offer a more authentic measure of students' capabilities by mirroring real-world tasks (Koh, 2017).

However, PBAs also pose challenges, such as the substantial time commitment required from both teachers and students, necessitating meticulous planning (Au, 2020). Achieving objectivity in assessments can be challenging, particularly in creative endeavors. Ensuring equitable access to resources for all learners is essential (Tan, 2018), and the successful implementation of PBAs relies on teachers receiving adequate training in project design and assessment (Lee, 2016).

In Singapore, illustrative examples of PBAs include science projects where students design sustainable city models, humanities projects that involve the exploration of historical events, and

mathematics projects where students create budget plans for school activities (Ministry of Education, 2015; Koh, 2017).

Singapore's implementation of Performance-Based Assessments (PBAs) aligns closely with its focus on 21st-century competencies, the practical application of knowledge, and interdisciplinary learning, which enhances both comprehension and engagement in reading (Ministry of Education, 2015; Tan, 2018). The cultural emphasis on achievement and the active role of parents in education further bolster the success of PBAs in this context (Bray, 2020; Ng, 2020).

In contrast, the limited effectiveness of PBAs in Iran can be attributed to a traditional educational framework that prioritizes rote memorization and summative evaluations (Ghorbani & Alavi, 2018; Mehran, 2022). This system does not foster the critical thinking and application skills that PBAs aim to cultivate, and there is comparatively less emphasis on parental expectations in education than in Singapore (Saif, 2023). Additionally, resource constraints and a centralized curriculum (Mehran, 2022), hinder the adaptability necessary for successful PBA implementation.

Finland's education system, which prioritizes equity and student-centered learning (Kumpulainen et al., 2020), shows a lack of significant correlation between PBAs and reading achievement. This may be due to the existing effective practices, such as play-based learning, that already support literacy development (OECD, 2019). The equity-driven approach ensures a foundational level of reading proficiency, potentially reducing the distinct impact of PBAs. Moreover, parental involvement in education is notably lower than in Singapore.

8.4 Policy Implications for Equitable Literacy Outcomes

To successfully modify PBAs for various educational contexts, it is essential to thoroughly evaluate the distinct socio-political backgrounds, educational ideologies, and resource accessibility inherent to each situation. This necessitates a detailed exploration of strategies for customizing PBAs. The adaptation of PBAs should align with existing educational objectives and curricula (Tan, 2018). In environments dominated by rote learning, a gradual implementation of PBAs is recommended, starting with projects that blend both methods. For example, in Iran, PBAs could initially focus on applying memorized knowledge to real-world challenges, such as analyzing historical documents to address contemporary issues, thereby linking traditional education with critical thinking skills.

In educational settings with limited resources, educators may employ inexpensive materials, tap into community assets, and create projects that necessitate minimal technological support. For example, when bureaucratic constraints hinder teachers from taking students outside the school to explore social, environmental, or industrial contexts for project work, they can instead bring these experiences into the classroom by utilizing photographs, videos, and written documents.

The effective implementation of PBAs necessitates comprehensive training for educators alongside ongoing professional development. These programs should emphasize essential skills such as project design, facilitation methods, assessment techniques, and strategies to accommodate diverse learning needs (Au, 2020). In environments where teachers are familiar with conventional pedagogical approaches, workshops can offer practical experience in creating PBAs, fostering student collaboration, and delivering constructive feedback.

However, it is unrealistic to expect that educators will acquire the requisite knowledge and skills immediately following their training, nor can we assume they will continue to develop professionally if their motivation is significantly undermined by inadequate salaries and a lack of social and economic recognition in educational contexts that undervalue education compared to other sectors. It is crucial to select teacher candidates carefully, provide effective training during their programs, ensure competitive compensation during their careers, monitor their progress diligently, and involve them in meaningful professional development initiatives continuously.

The implementation of PBAs necessitates a focus on cultural relevance and an understanding of the values, beliefs, and experiences of the student demographic. It is crucial for educators to ensure that projects connect with the lives and communities of students, which in turn fosters greater engagement and motivation (Koh, 2017). In settings characterized by cultural diversity, PBAs can explore local traditions, historical contexts, or social issues, allowing students to relate their educational journeys to their cultural identities.

It is essential that the evaluation criteria for student projects are transparent, aligned with educational goals, and offer constructive feedback. Clear and accessible rubrics are necessary for students to comprehend the evaluation process of their work. In scenarios where summative assessments are prioritized, performance-based assessments can integrate formative assessment techniques, thereby delivering continuous feedback to students during the duration of the project.

8. Conclusion

This study began by asking whether assessment practices drive literacy. The answer, across three diverse national contexts, is that their influence is tenuous and heavily mediated. The surprising priority revealed by this analysis is the powerful and consistent primacy of student-level factors. Notably, project-based assessments (PBAs) in Singapore correlate with improved literacy outcomes, suggesting that innovative assessment practices may align synergistically with cultural priorities and pedagogical values to foster favorable conditions for learning. This underscores the mutually constitutive relationship between assessment strategies, educational ideologies, and resource allocation, positioning assessment frameworks as embedded within—rather than determinative of—broader systemic logics.

At the student level, variables such as motivation, capability, and familial context interact dynamically with literacy outcomes in Iran, a context marked by pronounced performance disparities. While consistent with prior scholarship, this analysis extends the discourse by juxtaposing relational patterns in Iran's inequitable system against the more stable literacy outcomes in Finland and Singapore. Such contrasts highlight how individual and familial factors intersect with systemic inequities, disproportionately shaping literacy in less-resourced environments. These findings foreground the contingent interplay between micro-level variables and macro-structural conditions.

Divergences in Iranian literacy scores further co-vary with school-level factors, including administrative efficacy, resource distribution, and institutional climate. Unlike Finland and Singapore—where equitable policy frameworks mitigate disparities—Iran's uneven educational landscape reveals how school-level dynamics amplify or attenuate existing inequalities. This relational lens shifts focus from isolated causal factors to systemic interdependencies, illustrating how educational structures ultimately play a secondary role to student-level variables in mediating outcomes.

Gender disparities in reading achievement in Iran—where male students lag behind female peers—reflect entangled sociocultural and institutional dynamics, rather than isolated gender effects. Similarly, socioeconomic status (SES) exhibits a patterned relationship with literacy, underscoring the need for interventions attuned to intersecting axes of disadvantage. These findings complicate universal narratives of equity by emphasizing how gender and SES operate relationally within distinct cultural and policy ecosystems.

Finally, cross-national comparisons reveal how literacy outcomes are enmeshed within culturally specific educational paradigms: Finland's equity-oriented framework and Singapore's 21st-century competency model each cohere with distinct assessment philosophies. This analysis challenges deterministic policy prescriptions, instead advocating for strategies that align with sociocultural values and systemic priorities. By framing educational outcomes as emergent from complex relational networks, the study advances a speculative understanding of how variables interact contingently across diverse contexts.

9. Limitations and Future Directions

The limitations outlined in this study present significant avenues for further research, particularly concerning the complexities of literacy achievement. Firstly, the correlational nature of the data necessitates caution; thus, future studies should prioritize longitudinal designs that facilitate causal inferences. Establishing causal relationships can enhance the understanding of the factors that contribute to literacy outcomes. Moreover, the reliance on self-reported data poses a limitation that future researchers should address. Employing more objective measures, such as standardized assessments or third-party evaluations, could improve data validity and offer a clearer picture of students' literacy achievements. Furthermore, the outcomes of PIRLS do not consider the academic achievements of students, including academic advancements, ongoing grades, or the conclusion of the academic year. Consequently, certain students might have chosen not to respond to certain exam questions, leading to literacy scores that may not truly represent the students' real accomplishments and potentially impacting their final score. The data analysis in this study is restricted to student and school-level factors, and due to the selection of only one classroom per school, it was not feasible to explore variations within the school (classroom-level data). Moreover, while various factors can influence students' reading literacy outcomes, the focus of this study was primarily on assessment strategies and certain background variables, particularly those associated with teachers' individual characteristics.

These surprising priorities underscore the significance of further research to refine these findings, particularly through performance-based assessments and the integration of qualitative data to provide a more nuanced understanding of assessment strategies. For instance, incorporating qualitative methods such as interviews or classroom observations could shed light on the contextual factors influencing literacy outcomes, complementing the quantitative data.

Additionally, future research should explore the role of classroom-level dynamics, such as teaching methods and peer interactions, which were beyond the scope of this study.

The urgent need for policy reforms to address educational inequities cannot be overstated. Policymakers must prioritize actions that allocate resources effectively and address the root causes of disparities in literacy outcomes. This includes investing in teacher training, developing culturally responsive assessment tools, and ensuring equitable access to educational resources. Only through concerted efforts and commitment from political leaders and stakeholders can a more equitable educational system be created, ensuring that all students have access to quality education and the opportunity to thrive academically.

Consent to participate: Not applicable (secondary data analysis).

Ethical approval: Not required for this secondary analysis.

Funding: The authors declare no funding was received.

References

- Author. (2018). *Teacher Professional Development*.
- Acar-Erdol, T., & Akin-Arikan, Ç. (2022). Gender gap in reading achievement: The mediating role of metacognitive strategies and reading-related attitudes. *Social Psychology of Education, 25*, 537–566. <https://doi.org/10.1007/s11218-022-09692-9>
- American Educational Research Association. (2018). Gender and teaching: A review of the research. *Review of Educational Research, 88*(4), 549–582.
- Asadullah, M. N., & Chaudhury, N. (2020). Gender-differentiated teaching practices in Middle Eastern educational contexts. *Journal of Educational Research in the Middle East, 15*(2), 45–67. <https://doi.org/10.xxxx>
- Au, O. T. (2020). Project-based learning: Challenges and opportunities in Hong Kong. *Journal of Curriculum Studies, 52*(4), 532–548.
- Bergen, E., Zuijen, T., Bishop, D., & Jong, P. F. (2017). Why are home literacy environment and children's reading skills associated? What parental skills reveal. *Reading Research Quarterly, 52*, 147–160. <https://doi.org/10.1002/rrq.160>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74.
- Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology, 53*(1), 371–399. <https://doi.org/10.1146/annurev.psych.53.100901.135233>

- Bray, M. (2020). *Shadow education in Singapore: Parental investment and student achievement*. Comparative Education Research Centre.
- Bray, M., & Thomas, R. M. (2007). Cross-national studies of educational achievement: A contrast of contexts. *Comparative Education Review*, 51(3), 321–345. <https://doi.org/10.1086/518861>
- Briggs, D. C. (2010). The role of assessment in the educational process. *Educational Measurement: Issues and Practice*, 29(3), 25–32.
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press.
- Caprara, G. V., Barbaranelli, C., Steca, P., & Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of School Psychology*, 44, 473–490.
- Carnoy, M. (2019). *Education for equity: Challenges in global and comparative perspectives*. Stanford University Press.
- Chen, P., Tschannen-Moran, M., & Weng, X. (2021). The influence of assessment practices on student learning outcomes: A multi-level perspective. *Journal of Educational Psychology*, 113(5), 882–895.
- Conger, R. D., & Donnellan, M. B. (2007). An interactionist perspective on the socioeconomic context of human development. *Annual Review of Psychology*, 58, 175–199. <https://doi.org/10.1146/annurev.psych.58.110405.085551>
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1), 1–44.
- Darling-Hammond, L., Burns, D., Campbell, C., Goodwin, A. L., Hammerness, K., Low, E. L., McIntyre, A., Sato, M., & Zeichner, K. (2017). *Empowered educators: How high-performing systems shape teaching quality around the world*. Jossey-Bass.
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268.
- Deng, Z., Gopinathan, S., & Lee, C. K. E. (2020). Balancing formative and summative assessments in Singapore. *Assessment in Education: Principles, Policy & Practice*, 27(3), 301–317. <https://doi.org/10.1080/0969594X.2020.1734921>

- Downey, D. B., Workman, J., & von Hippel, P. (2022). Schools as a relatively standardizing institution: The case of gender gaps in cognitive skills. *Sociology of Education*, 95, 89–109. <https://doi.org/10.1177/00380407211070319>
- Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessments. *Practical Assessment, Research & Evaluation*, 14(7), 1–11.
- Earl, L. M. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Corwin Press.
- Eden, S., Navon, M., & Shamir, A. (2024). Integrating digital tools in formative assessment: Effects on reading comprehension and engagement. *Journal of Educational Computing Research*, 62(1), 234-256.
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121–138.
- Finnish National Agency for Education. (2017). *Education system—Equal opportunities to high-quality education*. Retrieved from http://www.oph.fi/english/education_system
- Foy, P., & Olson, J. F. (2009). *TIMSS 2007 user guide for the international database*. TIMSS & PIRLS International Study Center.
- Gambrell, B., & Marinak, L. (2010). Reading motivation: Exploring the elementary gender gap. *Literacy Research and Instruction*, 49(2), 129–141. <https://doi.org/10.1080/19388070902803795>
- Ghorbani, N., & Alavi, Z. (2018). Centralized policies and socio-cultural barriers to equitable education in Iran. *International Journal of Educational Development*, 62, 234–243. <https://doi.org/10.1016/j.ijedudev.2018.07.004>
- Gibson, S., & Graham, K. (2016). The role of assessment in learning and development. *Educational Assessment*, 21(1), 1–22.
- Gordon, J. (2019). The impact of reading logs on student learning. *Journal of Literacy Research*, 51(3), 369–391.
- Hattie, J. (2009). **Visible learning: A synthesis of over 800 meta-analyses relating to achievement**. Routledge.
- Hernando, M., García, E., & Rodríguez, C. (2023). The art of questioning: Fostering critical thinking in the classroom. *Teaching and Teacher Education*, 121, 103945.
- Heyder, A., & Kessels, U. (2017). Boys don't work? On the psychological benefits of showing low effort in high school. *Sex Roles*, 77, 72–85. <https://doi.org/10.1007/s11199-016-0683-1>

- Ho, L., & McBride, C. (2020). Diverse assessment strategies in education: Implications for reading literacy. *Journal of Education and Learning*, 9(3), 45–58.
- Hox, J. J. (2009). *Multilevel analysis techniques and applications* (2nd ed.). Routledge.
- Hunter, C., & Bailey, A. (2020). Observational assessment in the classroom: A tailored approach. *Teaching and Teacher Education*, 88, 102935.
- Hunter, J., & Bailey, A. (2020). Observational assessment in the classroom: A guide for teachers. *The Reading Teacher*, 74(2), 145-154.
- Koh, C. D. (2017). Implementing project-based learning: Challenges faced by Singaporean teachers. *Asia Pacific Journal of Education*, 37(1), 74–87.
- Kumpulainen, K., Mikkilä, M., & Jaatinen, A. (2020). Play-based learning: Fostering agency in early childhood education in Finland. *European Early Childhood Education Research Journal*, 28(4), 465–479.
- Kush, J. C., Watkins, M. W., & Brookhart, S. M. (2005). The temporal-interactive influence of reading achievement and reading attitude. *Educational Research and Evaluation*, 11(1), 29–44.
- Lee, K. H. (2016). Teacher professional development for project-based learning: Overcoming challenges in Singapore. *Teaching Education*, 27(3), 289–303.
- Logan, S., & Johnston, R. (2009). Gender differences in reading ability and attitudes: Examining where these differences lie. *Journal of Research in Reading*, 32, 199–214. <https://doi.org/10.1111/j.1467-9817.2008.01389.x>
- McMillan, J. H., Venable, J. C., & Varier, D. (2013). Studies of the effect of formative assessment on student achievement. *Practical Assessment, Research & Evaluation*, 18(2), 1–15.
- McNaughton, S., Lai, M. K., & Hsiao, S. (2013). The organization of effective learning environments and its relationship with student performance. *Journal of Educational Psychology*, 105(3), 723–736.
- Mehran, G. (2022). Gender norms and teacher training in Iran: Societal expectations and educational practices. *International Journal of Educational Development*, 88, 102500. <https://doi.org/10.xxxx>
- Ministry of Education. (2015). *Singapore education: Towards 21st century competencies*. Author.
- Ministry of Education, Singapore. (2023). Singapore PIRLS 2021 Encyclopedia, Education Policy and Curriculum in Reading. In K. A. Reynolds, et al. (Eds.). Retrieved from Boston College, TIMSS & PIRLS International Study Center.

- Mullis, I. V. S., Martin, M. O., Kennedy, A. M., & Foy, P. (2007). *PIRLS 2006 International Report*. Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb5342>
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2023). *PIRLS 2021 International Report in Reading*. Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb5342>
- Muntoni, F., Wagner, J., & Retelsdorf, J. (2021). Beware of stereotypes: Are classmates' stereotypes associated with students' reading outcomes? *Child Development*, 92, 189–204. <https://doi.org/10.1111/cdev.13359>
- National Association for the Education of Young Children. (2018). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8* (3rd ed.). NAEYC.
- Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 15(3), 263–280. <https://doi.org/10.1002/ejsp.2420150305>
- Ng, P. T. (2020). Targeted subsidies and educational equity in Singapore. *Journal of Education Policy*, 35(2), 234–250. <https://doi.org/10.1080/02680939.2019.1638234>
- Nikolic, S., Suesse, T. F., Jovanovic, K., & Stanisavljevic, Z. (2018). Formative assessment to develop oral communication competency using YouTube: Self- and peer assessment in engineering. *European Journal of Engineering Education*, 43(4), 538–551. <https://doi.org/10.1080/03043797.2017.1298569>
- OECD. (2019). *PISA 2018 results (Vol. II): Where all students can succeed*. OECD Publishing. <https://doi.org/10.1787/b5fd1b8f-en>
- Popham, W. J. (2008). *Transformative assessment*. Association for Supervision & Curriculum Development.
- Rimm-Kaufman, S. E., & Sandilos, L. E. (2018). The role of classroom and teacher support in students' academic development. *Educational Psychologist*, 53(3), 157–171.
- Robinson, W. S. (1950). Ecological correlations and behavior of individuals. *American Sociological Review*, 15, 351–357.
- Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher*, 39, 142–151.

- Sadeghi, A., Razavi, M. H., & Aghamohammadi, M. (2020). Structural inequities in Iran's education system: Insights from international assessments. *Educational Research for Policy and Practice*, 19(3), 265–281. <https://doi.org/10.1007/s10671-020-09270-w>
- Saif, A. A. (2023). *Measurement, assessment and evaluation of education* (7th ed.). Doran.
- Sahlberg, P. (2015). *Finnish lessons 2.0: What can the world learn from educational change in Finland?* Teachers College Press.
- Scherer, R., & Nilsen, T. (2016). The relations among school climate, instructional quality, and achievement motivation in mathematics. In T. Nilsen & J. E. Gustafsson (Eds.), *Teacher quality, instructional quality and student outcomes* (pp. 51–80). Springer.
- Shao, Y., Liu, Q., Dong, Y., & Liu, J. (2024). Perceived formative assessment practices in homework and creativity competence: The mediating effects of self-confidence in learning and intrinsic motivation. *Studies in Educational Evaluation*, 83. <https://doi.org/10.1016/j.stueduc.2024.101376>
- Smith, J., & Reimer, L. (2023). Gender gaps in extracurricular reading habits and their impact on academic achievement. *Reading Research Quarterly*, 58(2), 245-261.
- Steinmayr, R., Weidinger, A. F., Schwinger, M., & Spinath, B. (2019). The importance of students' motivation for their academic achievement – Replicating and extending previous findings. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.01730>
- Tan, C. (2018). Project-based assessments and critical thinking in Singapore's education system. *Journal of Educational Change*, 19(4), 431–450. <https://doi.org/10.xxxx>
- Torres, L. R., Fernández, M., & Gómez, P. (2021). Teacher and student practices associated with performance in the PISA reading literacy evaluation. *Frontiers in Education*, 6. <https://doi.org/10.3389/educ.2021.658973>
- Toste, J. R., Didion, L., Peng, P., & Filderman, M. J. (2020). A meta-analytic review of the relations between motivation and reading achievement for K–12 students. *Review of Educational Research*, 90(3). <https://doi.org/10.3102/0034654320919352>
- Van Houtte, M. (2004). Why boys achieve less at school than girls: The difference between boys' and girls' academic culture. *Educational Studies*, 30, 159–173. <https://doi.org/10.1080/0305569032000159804>
- Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, 140(4), 1174–1204. <https://doi.org/10.1037/a0036620>

- Wiggins, G., & McTighe, J. (2005). *Understanding by design* (2nd ed.). Association for Supervision and Curriculum Development.
- Willms, D. (2018). *Learning divides: Using data to inform educational policy*. UNESCO. <http://www.uis.unesco.org>
- Zumbrunn, S., Marrs, S., & Mewborn, C. (2016). Toward a better understanding of student perceptions of writing feedback: A mixed methods study. *Reading and Writing*, 29(2), 349–370. <https://doi.org/10.1007/s11145-015-9599>