Registered in Germany Europa Revue Journal
ISSN: 1862-6564 Vol. 6 No. 4 (2026) Published April 30, 2026

International Large-Scale Assessments as a Resource for External Quality Monitoring: Evidence from Uzbekistan's Participation in PISA, PIRLS, TIMSS and TALIS

Furkat G. Kholmirzayev

Republican Scientific and Methodological Centre for the Development of Education, Tashkent, Uzbekistan; A. Avloniy National Institute of Pedagogical Excellence, Tashkent, Uzbekistan

ORCID: https://orcid.org/0009-0007-0719-533X

Correspondence: furkatkholmirzayev@gmail.com

Abstract

Since 2021, Uzbekistan has taken part in four major international large-scale assessments (ILSAs): PIRLS 2021, PISA 2022, TIMSS 2023 and TALIS 2024, followed by PISA 2025. This marks a shift from a compliance-oriented tradition of quality control towards evidence-based external monitoring. This article examines what these studies reveal about the student-, family-, school- and teacher-level conditions of learning in Uzbekistan and how their evidence can be integrated into a national system of external monitoring of education quality. The study uses a secondary analysis of officially published international and national ILSA reports, combined with document analysis and a thematic synthesis organised around an input–process–outcome framework and the concept of educational equity. The results show that achievement in Uzbekistan is associated with the duration of pre-primary education, regular school attendance, school climate and students' educational expectations. Socio-economic gradients in achievement are markedly shallower than international averages: in PIRLS 2021 the reading gap between high- and low-SES pupils was 53 score points, against 86 points internationally. This pattern, however, must be read against low overall performance and the data-integrity concerns flagged in PISA 2025. TALIS 2024 shows that 87% of teachers feel their profession is valued by society (OECD average: 22%), yet their participation in school-level decision-making is between two and nine times lower than the OECD average, and principals report greater shortages of qualified staff, digital resources and library materials. The article argues that ILSAs yield their greatest value

when embedded in a coherent national monitoring architecture that safeguards data integrity, tracks equity indicators and engages teachers as active participants rather than implementers. A three-layer model of such an architecture and seven policy recommendations are proposed.

Keywords

international large-scale assessments; external monitoring; education quality; educational equity; PISA; PIRLS; TIMSS; TALIS; Uzbekistan; Central Asia

1. Introduction

Over the past quarter-century, international large-scale assessments (ILSAs) have become one of the most visible instruments of global education governance. The number of education systems taking part in the OECD Programme for International Student Assessment (PISA) grew from 45 in 2000 to 80 in 2022 (Turner & Li, 2026) and reached 91 in the 2025 cycle (OECD, 2026). Alongside PISA, the IEA studies PIRLS and TIMSS and the OECD Teaching and Learning International Survey (TALIS) now provide governments with comparable evidence on student achievement, the contexts of learning and the working conditions of teachers (Hastedt & Rocher, 2020). The rapid expansion of these programmes has generated an extensive literature on why countries participate and how the resulting data are used in national policy making (Addey et al., 2017; Fischman et al., 2019; Steiner-Khamsi & Waldow, 2018).

Uzbekistan is a recent but intensive participant in this movement. Within five years the country joined PIRLS (2021), PISA (2022), TIMSS (2023) and TALIS (2024), took part in PISA 2025 and entered the PIRLS 2026 cycle. This sequence of participation coincides with a broad national reform agenda that places explicit emphasis on the quality of education, the preparation of young people for independent life, the social status of teachers and the country's position in international rankings. Studying foreign experience, adopting international standards and modernising the national assessment system have therefore become priorities of education policy (Ismailov, 2022).

For a post-Soviet education system, this development has a significance that goes beyond the publication of league-table positions. Quality assurance in the region has historically relied on inspection, administrative reporting and the control of compliance

with centrally issued norms, whereas systematic, sample-based measurement of learning outcomes has been comparatively new (Kenzhetaeva et al., 2020; Rubtsova, 2023). ILSAs introduce not only a new source of data but also a new methodology of monitoring: probability sampling, standardised instruments, psychometric scaling and rich contextual questionnaires. They can thus serve as a model and a benchmark for the development of national external monitoring. At the same time, the comparative-education literature warns that ILSA results are frequently used selectively to legitimise reforms decided on other grounds (Fischman et al., 2019), to dramatise perceived failures ("scandalisation") or to project idealised images onto reference countries (Steiner-Khamsi & Waldow, 2018; Waldow, 2017). How a country learns to use ILSA evidence is therefore itself a question of policy importance.

Despite the scale of Uzbekistan's recent engagement, peer-reviewed English-language analyses that synthesise its results across several ILSAs remain scarce. National studies have concentrated mainly on adapting PISA methodology to the assessment of basic secondary education (Ismailov, 2022) and have not yet related student outcomes, family context and teacher conditions to one another. This article addresses that gap. Its aim is to analyse the evidence generated by Uzbekistan's participation in PISA, PIRLS, TIMSS and TALIS and to determine how this evidence can strengthen the national system of external monitoring of education quality. The study is guided by three research questions:

– RQ1. Which student-, family-, school- and teacher-level factors associated with learning outcomes in Uzbekistan are identified by PISA 2022, PIRLS 2021, TIMSS 2023 and TALIS 2024?

– RQ2. What do these studies indicate about educational equity in Uzbekistan, and how robust are such inferences?

– RQ3. How can ILSA evidence be integrated into the national architecture of external monitoring of education quality?

The article contributes to the literature in three ways. First, it provides the first cross-study synthesis of Uzbekistan's ILSA evidence, including the most recent PISA 2025 results. Second, it critically examines the apparent "equalising" pattern in Uzbekistan's

results rather than accepting it at face value. Third, it proposes a three-layer model for integrating international and national evidence into a coherent monitoring system, which may also be relevant to other Central Asian and post-Soviet systems that have recently joined ILSAs.

2. Literature Review and Conceptual Framework

2.1. ILSAs and national education policy

Early advocates of ILSAs emphasised their informational value: comparable data allow systems to benchmark their performance, set targets and identify policy levers. Breakspear (2012) showed that PISA had acquired a strong normative influence on national target setting and on the design of monitoring mechanisms in many OECD countries. Hastedt and Rocher (2020) further stress the potential of ILSAs to provide a reliable empirical basis for analysing educational equity, because their contextual questionnaires make it possible to compare outcomes across social groups and to identify differences in access to learning opportunities. From this perspective, ILSAs function as "thermometers" that diagnose the state of a system, although Lockheed and Wagemaker (2013) caution that they can easily turn into "whips" when rankings are used for political pressure rather than for improvement.

A second strand of research examines how ILSA data are actually used. Based on a literature review and surveys of policy experts, Fischman et al. (2019) found that ILSAs increasingly operate as "solutions in search of problems": they are used to legitimise reforms, while evidence of a direct causal link between participation and specific reforms remains limited; 38% of their respondents considered ILSAs to be generally misused in national policy contexts. Steiner-Khamsi and Waldow (2018) describe two typical national responses: scandalisation, in which comparison is used to highlight the weaknesses of one's own system, and projection, in which reference countries serve as screens onto which domestic ideals of "good" or "bad" education are projected. Gorur (2016) adds that ILSAs are performative: participation gradually leads policy makers to "see like PISA", adopting its categories as the natural language of quality.

Recent work shows that the design of an assessment shapes the evidence it produces and the policies it encourages. Cardoso (2022) demonstrated that differently designed

ILSAs, grounded in different institutional doctrines, can yield divergent estimates of the same policy effect, for example of grade repetition, and that participation in assessments is associated with changes in national repetition rates. Turner and Li (2026) model national participation in PISA using catastrophe theory and show that entering and leaving the programme are non-linear processes: once a country has invested politically in ILSA results, withdrawal becomes costly, and the path by which a country entered the programme affects how it responds to disappointing results. The Swedish case analysed by Ćosić (2024) illustrates the opposite dynamic: the decline observed in PISA 2012 triggered a sequence of system-wide reforms, showing how external assessment can become a major driver of national policy. Together, these studies suggest that the value of ILSAs for a participating country depends less on the data themselves than on the institutional arrangements through which the data are interpreted and used.

2.2. External monitoring of education quality

External monitoring refers to the systematic collection and analysis of evidence on the quality of education by bodies that are independent of the institutions being evaluated. Research on quality assurance consistently identifies a tension between its accountability and improvement functions. In a five-year study of external quality monitoring in a higher education college, Newton (1999) found that externally imposed procedures often generated compliance behaviour and bureaucratic burden rather than genuine improvement, and that the effect of monitoring depended on the extent to which staff took ownership of quality processes. Hayden and McKinlay (2001), describing a coordinated system of external visits in postgraduate medical education, showed that monitoring became more effective when standards were published in advance, institutions evaluated themselves against those standards between visits, evidence was triangulated from several sources and feedback shifted from a summative to a predominantly formative orientation. Similar conclusions emerge from school systems: in Punjab, Mahmood (2015) found internal monitoring to be more effective than external monitoring on most indicators, with external monitoring being more effective mainly in controlling teacher absenteeism, and recommended that external monitors provide pedagogical support rather than inspection alone.

The capacity of educators to engage with monitoring data is a recurring bottleneck. In a study of 112 university teachers in Russia, Rubtsova (2023) found that 72% experienced difficulties in conducting quality monitoring and 88% were not prepared to analyse learning outcomes, although two-thirds recognised the need to manage the quality of learning. In neighbouring Kazakhstan, the introduction of criteria-based assessment from 2016, itself strongly influenced by PIRLS, TIMSS and PISA, revealed that designing valid assessment criteria was the main difficulty for teachers (49% of respondents) and that parents' limited understanding of the new system hindered its acceptance (Kenzhetaeva et al., 2020). Measurement research also emphasises that large-scale and classroom assessment serve different but complementary purposes: the validity of classroom assessment depends above all on how well it serves teaching and learning (Kane & Wools, 2020), psychometric standards need to be translated into terms that are meaningful to teachers (Ferrara et al., 2020), and teacher-assigned grades capture aspects of student engagement that standardised tests do not (Bowers, 2020). A national monitoring system that relies exclusively on external tests therefore risks neglecting the classroom level at which improvement actually occurs.

2.3. Conceptual framework

This study combines an input–process–outcome perspective on education quality with the equity-indicator framework developed by the National Academies of Sciences, Engineering, and Medicine (2019). The latter distinguishes two types of indicators that a monitoring system should contain: indicators of disparities in student outcomes (readiness for school, learning and engagement, attainment) and indicators of disparities in access to opportunities and resources (early learning, high-quality curricula and instruction, and supportive school environments). The framework also draws a useful conceptual distinction between a between-group difference, which is a neutral numerical observation, and a disparity, which is a difference judged to matter for educational outcomes, and between equality (identical distribution of resources) and equity (distribution calibrated to need). An effective indicator system, according to this framework, should measure multiple dimensions, focus on policy-relevant subgroups,

be comparable across time and place, and be based on scientifically sound measures. Table 2 in Section 3 maps these domains onto the ILSA data analysed in this article.

3. Methodology

3.1. Research design

The study adopts a qualitative-dominant design based on secondary analysis of published quantitative data, complemented by document analysis (Cohen et al., 2013; Creswell, 2014). Secondary analysis of ILSA reports is appropriate for the research questions because these reports are produced under strict international technical standards (probability sampling, translation verification, psychometric scaling and weighting), which makes their national estimates more reliable than data that could be collected in a small-scale local survey. The design is descriptive and comparative: national estimates for Uzbekistan are compared with international or OECD averages and across subgroups within the country.

3.2. Data sources

The primary data sources are the international reports and national analytical materials of the four ILSAs in which Uzbekistan participated between 2021 and 2025 (Table 1). Student-level results were drawn from PISA 2022 (OECD, 2023a, 2023b), PIRLS 2021 (Mullis et al., 2023) and TIMSS 2023 (von Davier et al., 2024); principal- and teacher-level results were drawn from the contextual questionnaires of TIMSS 2023 and TALIS 2024 (OECD, 2025). The first volume of PISA 2025 results (OECD, 2026) was used to update the analysis. The assessment frameworks of these studies (Mullis et al., 2021; OECD, 2018) were used to interpret the constructs measured. In addition, national analytical reports on Uzbekistan's participation prepared by the education authorities were analysed as policy documents.

Table 1. International large-scale assessments analysed in the study

Study	Organiser	Target population	Main domains	Cycle(s) with Uzbekistan
PIRLS	IEA	Grade 4 students	Reading literacy	2021 (first); 2026 (ongoing)
PISA	OECD	15-year-old students	Reading, mathematics and science literacy	2022 (first); 2025
TIMSS	IEA	Grade 4 students	Mathematics and science	2023 (first)
TALIS	OECD	Lower-secondary teachers and principals	Teaching conditions, professional development, school leadership	2024 (first)

Source: compiled by the author from OECD (2023a, 2025, 2026), Mullis et al. (2023) and von Davier et al. (2024).

3.3. Analytical procedure

The analysis proceeded in three stages. First, national estimates relevant to the research questions, together with the corresponding international or OECD averages, were extracted from the reports and organised in analytical tables and figures. Second, the extracted indicators were coded according to the level at which they operate (student, family, school, teacher, system) and mapped onto the equity-indicator domains of the National Academies of Sciences, Engineering, and Medicine (2019), as shown in Table 2. Third, findings were synthesised thematically across studies, with particular attention to patterns that were consistent across different assessments. Triangulation across four independently designed studies served as the main strategy for strengthening the credibility of the conclusions, following the principle that information from a single

source is less reliable than convergent evidence from several (Hayden & McKinlay, 2001).

Table 2. Analytical framework: levels, indicator domains and ILSA sources

Level	**Indicator domain (NASEM, 2019)**	**Source**	**Indicators examined**
Student	K–12 learning and engagement	PISA 2022	Attendance, homework time, educational expectations, life satisfaction
Family	Contextual and structural factors	PIRLS 2021; TIMSS 2023	Socio-economic status, home resources, parents' enjoyment of reading, school socio-economic composition
Early learning	Access to high-quality early learning	PISA 2022	Duration of pre-primary education
School	Supportive school environments	PISA 2022; TALIS 2024	School climate index, principals' reports of resource shortages
Teacher	Access to high-quality instruction	TALIS 2024	Perceived status, barriers to professional development, participation in decision-making
System	Quality and comparability of evidence	PISA 2025	Reporting notes on data quality and comparability

Source: developed by the author on the basis of National Academies of Sciences, Engineering, and Medicine (2019).

3.4. Ethical considerations

The study relies exclusively on publicly available, aggregated and anonymised data published by international organisations and national authorities. No personal data were collected, and no human participants were involved; ethical approval was therefore not required.

4. Results

4.1. Student- and school-level factors: evidence from PISA 2022

In its first PISA cycle, Uzbekistan recorded mean scores of 364 points in mathematics, 336 in reading and 355 in science, well below the OECD averages of 472, 476 and 485 respectively (OECD, 2023a). Beyond these headline figures, the PISA contextual data show that students' achievement is shaped not only by the content of instruction but by a combination of factors that include time for learning, school environment, attitudes towards school and expectations for the future.

Time-related resources for learning are associated with performance in mathematics (Figure 1). Students who attended pre-primary education, who had not been absent from school for more than three consecutive months, and who spent regular time on homework achieved higher results. Conversely, prolonged absence and insufficient home study were accompanied by a substantial decline in achievement. When the socio-economic status of students and schools is taken into account, the size of these associations changes, which indicates that part of the observed differences is linked to family background, but the associations do not disappear.

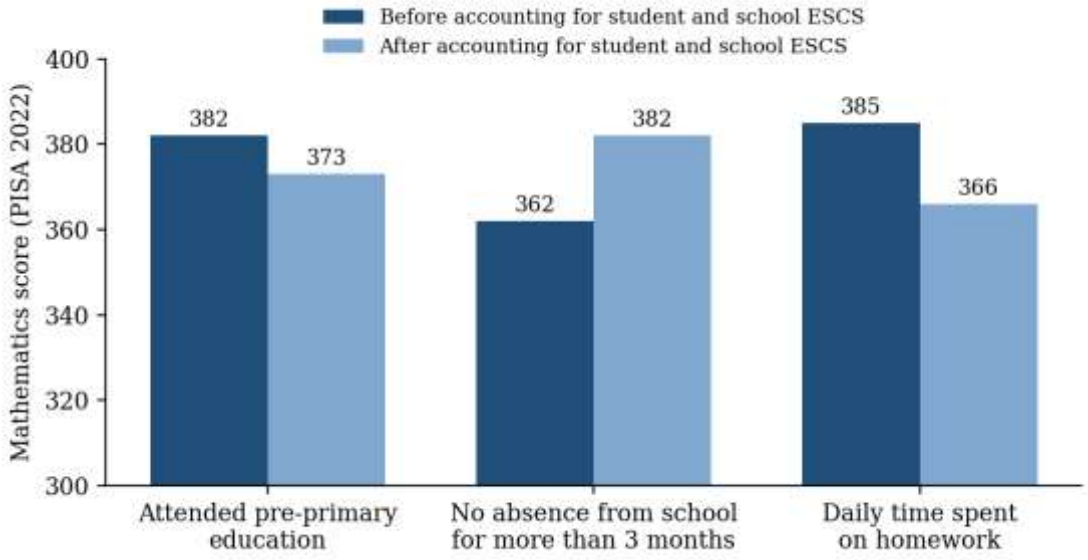

Figure 1. Time-related resources for learning and mathematics performance of students in Uzbekistan, PISA 2022

Note: ESCS = PISA index of economic, social and cultural status. Source: author's compilation based on OECD (2023b) and national analysis of PISA 2022 data.

A positive school climate is also associated with better results. Students in the top quarter of the index of student-related factors affecting school climate scored 385 points in mathematics, compared with 378 points in the bottom quarter (Figure 2). The difference is modest, but it is consistent with the broader finding that students who feel safe, respected and supported at school tend to hold more positive attitudes towards learning. Students' sense of belonging operates through similar mechanisms: students who perceive themselves as full members of the school community show higher motivation and discipline, whereas those who feel alienated from school display lower interest in learning and lower achievement (OECD, 2023b).

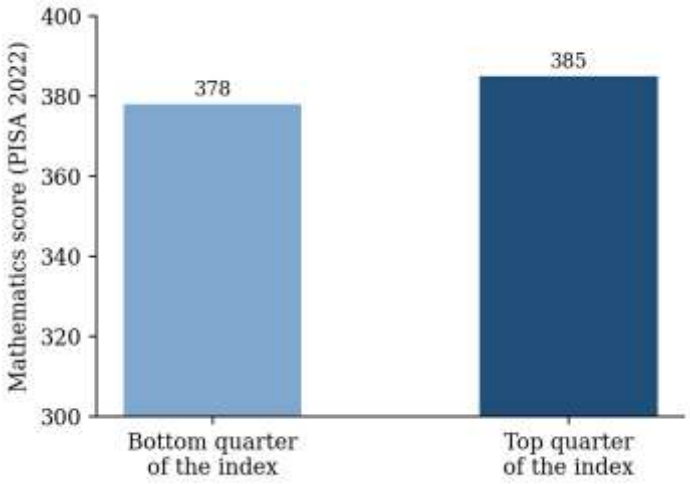

Figure 2. Index of student-related factors affecting school climate and mathematics performance, Uzbekistan, PISA 2022

Source: author's compilation based on OECD (2023b).

The duration of pre-primary education shows one of the clearest gradients. Students who attended pre-primary education for three years or more outperformed those who attended for one to less than two years by 19 points in mathematics, 26 points in reading and 19 points in science (Figure 3). This finding is particularly relevant for Uzbekistan, where the expansion of pre-school enrolment has been a major policy priority in recent years.

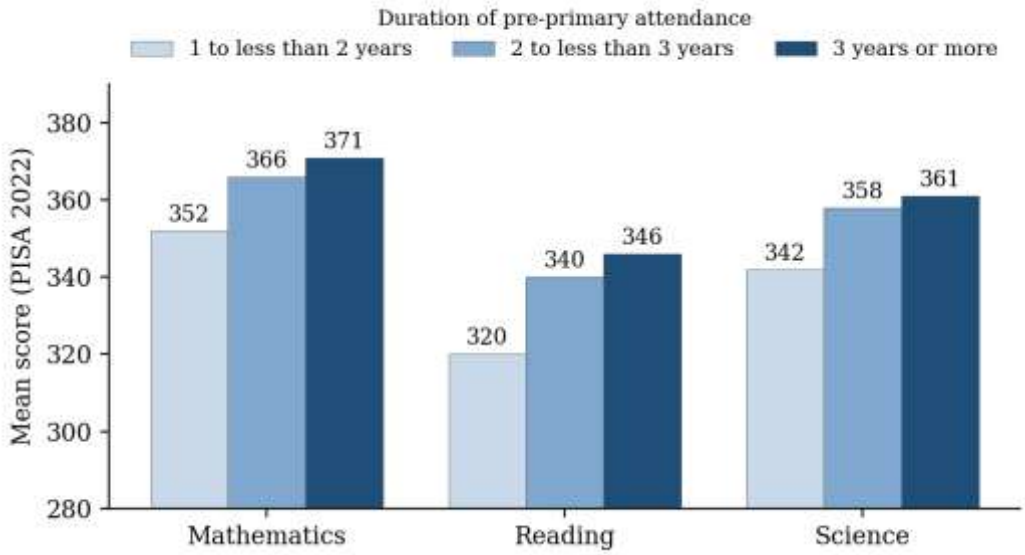

Figure 3. Duration of pre-primary education and performance in mathematics, reading and science, Uzbekistan, PISA 2022

Source: author's compilation based on OECD (2023b).

Students' educational expectations are also related to achievement. Among high-performing students, 76% expected to complete tertiary education, compared with 68% among low-performing students, while the share expecting to leave education below upper secondary level was more than twice as high among low performers (15% versus 6%) (Figure 4). Students with ambitious educational goals tend to invest more effort in learning, whereas low aspirations are usually accompanied by lower achievement.

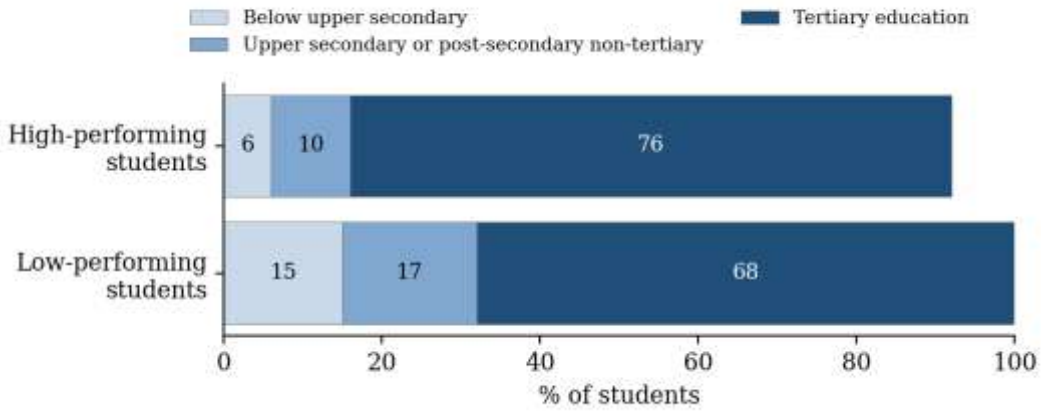

Figure 4. Students' expected level of educational attainment, by mathematics performance, Uzbekistan, PISA 2022

Note: percentages may not add up to 100 because of missing responses. Source: author's compilation based on OECD (2023a).

PISA also examines students' socio-emotional well-being. According to the national analysis of PISA 2022 questionnaire data, around 90% of 15-year-old students in Uzbekistan reported being satisfied with their lives. While this is an encouraging indicator, it cannot by itself be interpreted as an effect of recent reforms, since life satisfaction is influenced by many factors outside school. Internationally, PISA 2022 also found that in education systems where parental involvement remained stable or declined less between 2018 and 2022, mathematics performance tended to be stable or to improve, which underlines the importance of school–family engagement (OECD, 2023b).

4.2. Family context and reading: evidence from PIRLS 2021

PIRLS assesses the reading literacy of fourth-grade students every five years. Grade 4 is chosen because by this stage children have usually acquired basic reading skills and

begin to use reading as a tool for learning other subjects (Mullis et al., 2023). The PIRLS 2021 data allow the relationship between reading achievement, family socio-economic status and parents' attitudes towards reading to be examined.

Across participating countries, students from high-SES families scored on average 543 points, compared with 457 points for students from low-SES families, a gap of 86 points. In Uzbekistan the corresponding scores were 474 and 421 points, a gap of 53 points (Figure 5). Socio-economic advantage is thus associated with higher reading achievement in Uzbekistan as elsewhere, confirming the importance of home resources, parental education and access to books. The gap, however, is markedly smaller than the international average, which suggests that school may partly perform a compensatory function by providing pupils from different social groups with more comparable opportunities to develop reading skills. The robustness of this interpretation is discussed in Section 5.2.

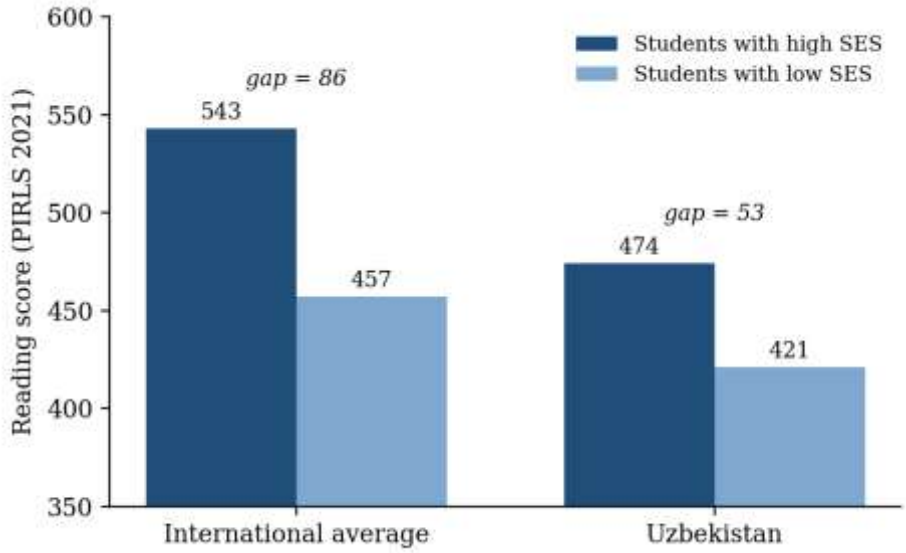

Figure 5. Reading achievement of students with high and low socio-economic status: international average and Uzbekistan, PIRLS 2021

Note: the PIRLS home socio-economic status scale is based on parents' reports of home resources, parental education and occupation. Source: author's compilation based on Mullis et al. (2023).

Parents' attitudes towards reading are closely related to students' outcomes. In Uzbekistan, 39% of parents reported that they very much like reading, while 4% reported

that they do not like reading. The children of parents who very much like reading scored 449 points, compared with 431 points for children of parents who somewhat like reading and 404 points for children of parents who do not like reading (Figure 6). The 45-point difference between the extreme groups indicates that the family's reading culture is one of the key factors in children's literacy development.

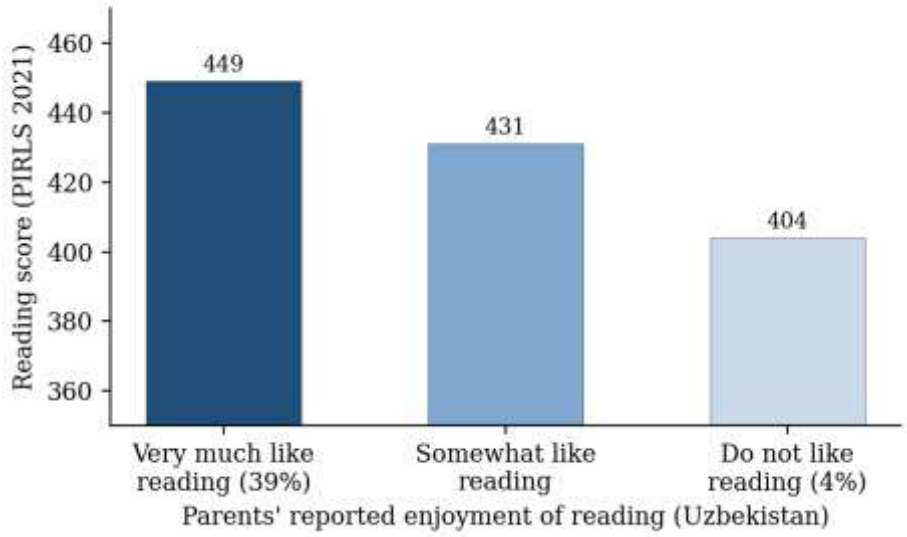

Figure 6. Parents' enjoyment of reading and students' reading achievement, Uzbekistan, PIRLS 2021

Source: author's compilation based on Mullis et al. (2023).

4.3. School socio-economic composition: evidence from TIMSS 2023

TIMSS assesses the mathematics and science knowledge of students every four years, with an emphasis on core subject knowledge and on the ability to apply it and reason when solving problems (Mullis et al., 2021). According to principals' reports in TIMSS 2023, 56% of fourth-grade students in Uzbekistan attended schools with more affluent than disadvantaged students, 39% attended schools with neither composition predominating, and only 4% attended schools with more disadvantaged than affluent students. Most notably, average achievement barely differed across these groups: mathematics scores ranged from 443 to 447 points and science scores from 410 to 413 points (Table 3).

Table 3. Grade 4 achievement by school socio-economic composition (principals' reports), Uzbekistan, TIMSS 2023

School composition by students' economic background	% of students	Mathematics	Science
More affluent than disadvantaged	56	444	411
Neither more affluent nor more disadvantaged	39	443	413
More disadvantaged than affluent	4	447	410

Source: author's compilation based on von Davier et al. (2024).

Taken at face value, these data suggest that success in mathematics and science is linked less to the economic composition of the school than to the quality of the teaching process, the professional work of teachers and students' own motivation. This pattern is consistent with the principle of equal educational opportunity. However, as discussed below, a school-level composition measure based on principals' estimates is less sensitive than individual-level socio-economic measures, and the "disadvantaged" group represents only 4% of students, which limits the precision of the comparison.

4.4. Teachers and school resources: evidence from TALIS 2024

As Schleicher (2018) repeatedly emphasises, the quality of an education system cannot exceed the quality of its teachers. Sustained improvements in education quality have been achieved primarily in systems where teachers are highly competent, have ample opportunities for continuous professional development, participate in school governance and enjoy recognised social status. In 2024 teachers in Uzbekistan took part in TALIS for the first time, which made it possible to analyse their work, professional development and working conditions in international comparison.

One of the most striking findings concerns the perceived status of the profession: 87% of teachers in Uzbekistan reported that they feel their profession is valued in society, compared with an OECD average of 22%, placing Uzbekistan among the systems with the highest values on this indicator (OECD, 2025). This result is consistent with a series

of recent government measures aimed at raising teachers' social status and providing material and moral incentives.

With regard to professional development, teachers in Uzbekistan report lower barriers than their OECD colleagues on most dimensions, including conflicts with the work schedule, lack of time and lack of incentives (Table 4). Two exceptions are important. Lack of employer support is reported at almost the same level as the OECD average, and the lack of digital infrastructure is reported twice as often as in the OECD (27.9% versus 13.8%). Digital infrastructure therefore emerges as the specific bottleneck for professional learning in Uzbekistan, which is particularly relevant given the growing role of online and blended professional development.

Table 4. Barriers to teachers' participation in professional development, Uzbekistan and OECD average, TALIS 2024 (% of teachers)

Barrier	Uzbekistan	OECD average	Difference (pp)
Conflict with work schedule	22.4	58.7	−36.3
Lack of time due to other responsibilities	28.1	63.4	−35.3
Lack of employer support	27.5	26.9	+0.6
Lack of incentives for participation	34.3	45.7	−11.4
Distance and geographical factors	26.5	31.3	−4.8
Lack of digital infrastructure	27.9	13.8	+14.1

Note: pp = percentage points. Source: author's compilation based on OECD (2025).

Principals' reports indicate that shortages of several key resources are more acute in Uzbekistan than on average in the OECD (Table 5). The shortage of qualified teachers was reported as hindering the quality of instruction by 33.7% of principals, compared with 23.1% in the OECD. The shortages of digital resources and equipment (29.0% versus 16.6%) and of library materials (25.0% versus 11.0%) are almost twice as

frequent, and the shortage of buildings and instructional space is also more common (33.1% versus 25.0%).

Table 5. Principals reporting that resource shortages hinder the school's capacity to provide quality instruction, TALIS 2024 (%)

Resource	Uzbekistan	OECD average	Ratio
Qualified teachers	33.7	23.1	1.5
Digital technology and equipment	29.0	16.6	1.7
Library materials	25.0	11.0	2.3
Buildings and instructional space	33.1	25.0	1.3

Note: ratio = Uzbekistan value divided by OECD average. Source: author's compilation based on OECD (2025).

School-level policies on discipline, assessment, inclusive education and professional development are more sustainable and better aligned with real needs when they are shaped with the participation of teachers. TALIS 2024 shows, however, that the involvement of teachers in Uzbekistan in such decisions remains low (Table 6). Only 19.6% of teachers reported participating in determining school discipline policy (OECD average: 39.7%), 16.4% in setting school development priorities (34.3%), 10.9% in shaping professional development policy (40.2%), 7.7% in determining inclusive education policy (39.7%) and just 2.2% in policies related to staff diversity (20.2%). In other words, on these issues teachers act mainly as implementers of decisions already taken.

Table 6. Teachers reporting participation in school-level policy decisions, TALIS 2024 (%)

Policy area	Uzbekistan	OECD average	Ratio (OECD / Uzbekistan)
School discipline policy	19.6	39.7	2.0

Policy area	Uzbekistan	OECD average	Ratio (OECD / Uzbekistan)
School development priorities	16.4	34.3	2.1
Professional development policy	10.9	40.2	3.7
Inclusive education policy	7.7	39.7	5.2
Staff diversity policy	2.2	20.2	9.2

Source: author's compilation based on OECD (2025).

4.5. Update: PISA 2025

The first volume of PISA 2025 results, published in September 2026, adds two important observations. First, Uzbekistan's mean science score rose to 438 points, 83 points higher than in 2022 and one of the largest increases among participating systems, although still below the OECD average of 482 points; 65% of students reached at least the baseline proficiency Level 2 in science (OECD average: 74%). Socio-economic status accounted for only about 1% of the variation in science performance, compared with 12% on average across the OECD (OECD, 2026). Second, the OECD identified a number of inconsistencies in the mathematics and reading response data from Uzbekistan that required further investigation and appeared to be more frequent in some schools and regions. Because these inconsistencies undermine the comparability of scale scores, Uzbekistan's results in mathematics and reading were not reported, and only science, computational problem-solving and questionnaire-based results were published (OECD, 2026).

5. Discussion

5.1. Conditions of learning in Uzbekistan (RQ1)

Taken together, the four studies present a consistent picture. At the student level, achievement is associated with time for learning (regular attendance, homework and, especially, the duration of pre-primary education), with a supportive school climate and with students' educational aspirations. At the family level, home resources and, above all, a family culture of reading are strongly related to early literacy. At the school and

teacher level, the key constraints are not the motivation or the perceived status of teachers, which are exceptionally high, but the availability of qualified staff, digital infrastructure and library resources, and the limited involvement of teachers in school decision-making. These findings indicate that improving outcomes requires action beyond curriculum revision: developing the school environment, strengthening learning motivation, supporting family engagement and building effective daily learning habits are equally important.

The duration of pre-primary education deserves particular attention. The 19- to 26-point gradients observed in PISA are substantial relative to the country's mean scores and are consistent with international evidence on the long-term value of early learning, which the National Academies of Sciences, Engineering, and Medicine (2019) identify as one of the core domains of equitable access to opportunity. Because the cohort assessed in PISA 2022 entered pre-school before the recent large-scale expansion of pre-school provision, subsequent PISA cycles will provide an opportunity to monitor whether this expansion translates into improved outcomes at age 15.

5.2. An equalising system or a compressed distribution? (RQ2)

The most policy-relevant and, at the same time, the most delicate finding concerns equity. All three student assessments point in the same direction: the PIRLS reading gap between high- and low-SES students is smaller than the international average, TIMSS results barely differ across schools of different socio-economic composition, and in PISA 2025 socio-economic status explained only about 1% of the variation in science performance. It is tempting to conclude that Uzbekistan's school system successfully neutralises social disadvantage. The literature on the policy uses of ILSAs, however, warns against such readings: favourable results can easily become the basis of a national self-image that is projected onto the data rather than derived from them (Steiner-Khamsi & Waldow, 2018; Waldow, 2017).

At least four alternative or complementary explanations need to be considered. First, when overall performance is low and a large share of students performs at similar, modest levels, the distribution of scores is compressed and between-group differences shrink regardless of the degree of equity; a small gap at a low level of achievement is

not equivalent to a small gap at a high level. Second, the measures of socio-economic status differ between studies: TIMSS uses principals' estimates of school composition, which are less discriminating than individual indices, and the most disadvantaged category comprises only 4% of students. Third, cross-national socio-economic indices may capture economic differences less accurately in contexts where household assets and parental occupations are distributed differently from the OECD countries for which the indices were originally developed. Fourth, the data-quality concerns identified in PISA 2025 show that the integrity of test administration must itself be treated as a condition for any valid inference about equity. In the terms of the National Academies of Sciences, Engineering, and Medicine (2019), the observed patterns are best described as between-group differences that are encouraging, but that require confirmation from national data with individual-level socio-economic measures before they can be interpreted as evidence of an equitable system.

5.3. Valued but not empowered: the teacher paradox

TALIS 2024 reveals a paradox. Teachers in Uzbekistan feel that their profession is valued by society to an extent rarely seen in other systems, yet they participate in school-level policy making far less than their OECD colleagues. High social recognition combined with limited professional agency is characteristic of centralised systems in which reforms are designed at the top and transmitted downwards. Research on quality monitoring suggests that such an arrangement limits the sustainability of improvement: externally driven procedures tend to produce compliance rather than ownership (Newton, 1999), whereas monitoring becomes effective when institutions and practitioners participate in evaluating themselves against shared standards (Hayden & McKinlay, 2001). The experience of Kazakhstan and Russia confirms that new assessment systems falter when teachers lack the preparation and involvement needed to apply them (Kenzhetaeva et al., 2020; Rubtsova, 2023). The very high proportion of teachers who report positive perceptions and the low reported barriers to professional development should also be interpreted with some caution, since self-reports in survey research may be influenced by social desirability and culturally specific response styles (Cohen et al., 2013).

5.4. Towards an integrated architecture of external monitoring (RQ3)

The Uzbek case supports the conclusion of Fischman et al. (2019) that participation in ILSAs does not automatically lead to evidence-based policy: the benefit depends on the institutional arrangements that link international data to national decision making. It also illustrates the insight of Turner and Li (2026) that the way a country engages with ILSAs shapes its future options. A country that uses ILSAs mainly to seek short-term political benefit becomes vulnerable to sudden and costly reversals when results disappoint, whereas a long-term orientation towards benchmarking and capacity building makes engagement more stable. Cardoso (2022) further shows that each assessment carries the assumptions of its design, which argues for triangulating evidence across several ILSAs and national sources rather than relying on any single ranking.

On this basis, the article proposes a three-layer model of an integrated national monitoring architecture (Figure 7). The first layer comprises the sources of evidence: ILSAs, national sample-based assessments, administrative data and school self-evaluation, including classroom assessment. The second, integration layer contains a common indicator framework covering outcomes, opportunities and equity; procedures for assuring data quality and the integrity of test administration; independent secondary analysis; and a regular public reporting cycle. The third layer concerns the use of evidence at the system, school, teacher and public levels, while a feedback loop ensures that the evaluation of reform effects informs subsequent assessment cycles and the revision of indicators.

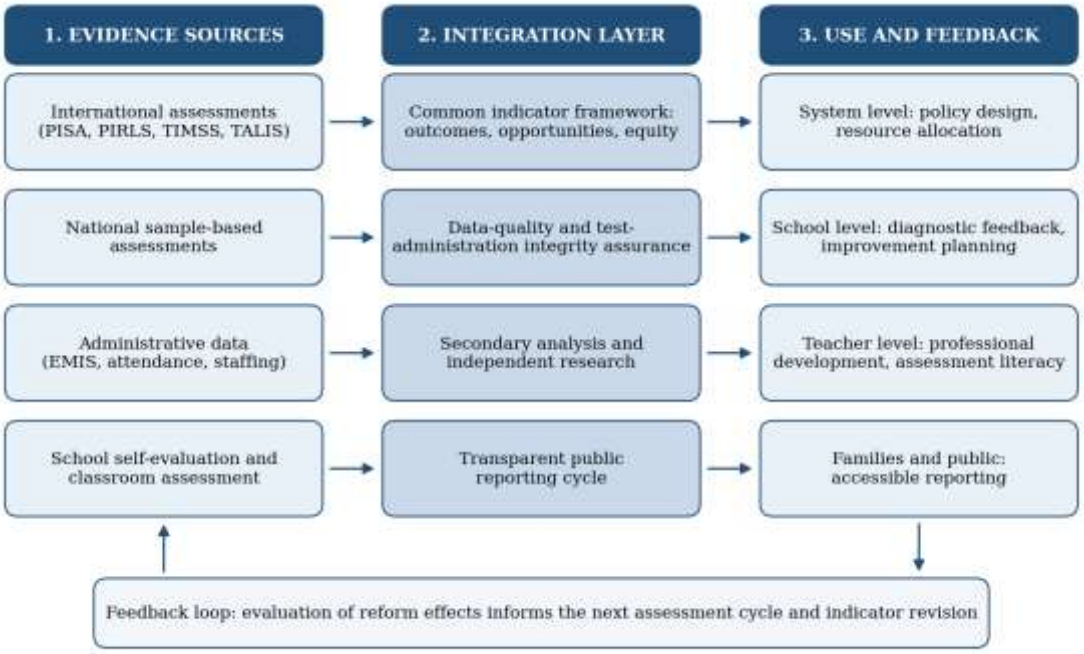

Figure 7. Proposed three-layer model of an integrated national architecture for external monitoring of education quality

Source: developed by the author.

Emerging technologies may support such an architecture but also raise new questions. Artificial intelligence is increasingly used to process multimodal learning data and to provide rapid feedback (Wang, 2026), and it can assist large-scale assessment programmes at every stage, from item development to secure administration and reporting, provided that human experts retain responsibility for defining and validating constructs (Lorié & Dadey, 2026). At the same time, generative AI challenges the credibility of learner performance and may reinforce surveillance and inequality (Broadfoot & Rockey, 2025). For Uzbekistan, the most immediate value of such tools may lie in strengthening the monitoring of test administration and the detection of anomalous response patterns, precisely the area highlighted by PISA 2025.

6. Policy Recommendations

Based on the findings, the following recommendations are proposed for the development of the national system of assessment and external monitoring of education quality:

1. **Integrate national assessments with international frameworks.** The methodological approaches of PISA, PIRLS and TIMSS, including probability

sampling, psychometric scaling and contextual questionnaires, should be incorporated into national sample-based assessments, while anchoring their content in the national curriculum and reflecting the relevant criteria in national assessment instruments.

2. **Safeguard the integrity of assessment data.** In light of PISA 2025, national standards for test administration, independent quality monitoring during data collection and routine statistical screening for anomalous response patterns should be established, and the investigation of PISA 2025 data should be completed transparently in cooperation with the OECD.

3. **Establish a national equity indicator set.** Achievement gaps by socio-economic status, region, urban–rural location, gender and language of instruction should be reported regularly, using individual-level socio-economic measures, so that the apparent equalising function of schools can be verified rather than assumed.

4. **Expand access to pre-primary education and monitor its quality.** Given the strong association between the duration of pre-primary attendance and later achievement, expansion should be accompanied by monitoring of quality and by tracking the outcomes of expanded cohorts in future ILSA cycles.

5. **Develop a supportive school environment and a family culture of reading.** Schools should be places where students feel safe, respected and involved. Systematic work with parents, including reading programmes that start in early childhood, should become a priority of primary education policy.

6. **Increase teachers' participation in school decision-making.** Teachers should be involved in shaping school policies on discipline, assessment, inclusive education and professional development, and their assessment literacy should be strengthened so that they can use monitoring data for improvement.

7. **Remove the digital bottleneck in professional development.** Investment in digital infrastructure and resources in schools, together with flexible and accessible formats of professional learning, should address the main barrier identified by TALIS 2024 as well as the shortages of qualified teachers and library materials reported by principals.

7. Limitations and Future Research

This study has several limitations. First, it relies on published aggregate estimates rather than on analysis of the international microdata; consequently, standard errors and the statistical significance of some differences could not be examined, and the results are descriptive and correlational rather than causal. Second, the four assessments differ in target populations, years and constructs, so the cross-study synthesis identifies convergent patterns rather than directly comparable estimates. Third, the self-reported data from teachers and principals may be affected by response styles. Fourth, the comparability of Uzbekistan's PISA 2025 mathematics and reading data is under review, which limits trend analysis. Future research should analyse the PISA, PIRLS, TIMSS and TALIS microdata using appropriate methods (plausible values, sampling weights and replicate weights), apply multilevel models to examine the equity patterns identified here, and link ILSA data to national assessment and administrative data to evaluate the effects of specific reforms.

8. Conclusion

Uzbekistan's active participation in PISA, PIRLS, TIMSS and TALIS has substantially expanded the possibilities for diagnosing the state of the national education system and evaluating it on the basis of comparable indicators. PISA results show a direct relationship between student achievement and factors such as time for learning, pre-primary education, school climate and students' expectations for the future. PIRLS and TIMSS show that outcomes are closely linked to the family environment and parents' attitudes to reading, while also suggesting that schools may partly perform a compensatory, equalising function, an interpretation that requires further verification. TALIS reveals that teachers are highly valued but insufficiently involved in school-level decisions and that professional development is constrained above all by digital infrastructure. The PISA 2025 experience demonstrates that the integrity of assessment data is itself a fundamental dimension of education quality.

The overall conclusion is that ILSAs are most valuable not as a source of rankings but as a methodological model and a source of evidence for a national monitoring system that combines international and national data, safeguards data integrity, tracks equity

and engages teachers as partners. Moving from participation in international studies to the systematic use of their evidence is the next and decisive stage in the development of external monitoring of education quality in Uzbekistan.

Declarations

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflict of interest: The author is employed by the Republican Scientific and Methodological Centre for the Development of Education and takes part, together with the A. Avloniy National Institute of Pedagogical Excellence (the national centre for PISA and PIRLS in Uzbekistan), in the national implementation of international large-scale assessments, including PISA, PIRLS and TIMSS. In particular, the author served as a test administrator in the PISA 2025 Main Survey (Order of the Minister of Preschool and School Education of the Republic of Uzbekistan No. 133 of 25 April 2025) and is involved in the organisation of the PIRLS 2026 Main Survey (Order No. 122 of 2 April 2026). The analysis presented in this article is based on publicly released international reports and national analytical materials; no confidential school- or student-level information collected during these studies was used. The views expressed are those of the author and do not necessarily reflect the official position of the Ministry of Preschool and School Education or of the institutions involved. The author declares no other competing interests.

Data availability: All data analysed in this study are publicly available in the sources listed in the References.

Ethics statement: The study used publicly available aggregated data only; no human participants were involved and ethical approval was not required.

Declaration of generative AI use: The research design, selection of sources, data analysis, interpretation of results and conclusions are the author's own work. The manuscript was originally written by the author in Russian. During its preparation for publication in English, the author used Claude (Anthropic) for translation, language editing and assistance with structuring the text in accordance with the journal's format.

The author reviewed, verified and edited all content, including all figures, data and references, and takes full responsibility for the integrity and content of the publication.

References

Addey, C., Sellar, S., Steiner-Khamsi, G., Lingard, B., & Verger, A. (2017). The rise of international large-scale assessments and rationales for participation. *Compare: A Journal of Comparative and International Education, 47*(3), 434–452. https://doi.org/10.1080/03057925.2017.1301399

Bowers, A. J. (2020). Towards measures of different and useful aspects of schooling: Why schools need both teacher-assigned grades and standardized assessments. In S. M. Brookhart & J. H. McMillan (Eds.), *Classroom assessment and educational measurement* (pp. 209–223). Routledge. https://doi.org/10.4324/9780429507533-12

Breakspear, S. (2012). *The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance* (OECD Education Working Papers No. 71). OECD Publishing. https://doi.org/10.1787/5k9fdfqffr28-en

Broadfoot, P., & Rockey, J. (2025). Generative AI and the social functions of educational assessment. *Oxford Review of Education, 51*(2), 281–300. https://doi.org/10.1080/03054985.2025.2455549

Cardoso, M. E. (2022). *Policy evidence by design: How international large-scale assessments influence repetition rates* [Doctoral dissertation, Columbia University]. Columbia Academic Commons.

Cohen, L., Manion, L., & Morrison, K. (2013). *Research methods in education* (7th ed.). Routledge. https://doi.org/10.4324/9780203720967

Ćosić, M. (2024). Interrelation between educational reforms and external assessment of student achievements in the context of the Swedish educational system. *Školski vjesnik, 73*(1), 19–33. https://doi.org/10.38003/sv.73.1.2

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE.

Ferrara, S., Maxey-Moore, K., & Brookhart, S. M. (2020). Guidance in the Standards for classroom assessment: Useful or irrelevant? In S. M. Brookhart & J. H. McMillan

(Eds.), *Classroom assessment and educational measurement* (pp. 97–119). Routledge. https://doi.org/10.4324/9780429507533-7

Fischman, G. E., Topper, A. M., Silova, I., Goebel, J., & Holloway, J. L. (2019). Examining the influence of international large-scale assessments on national education policies. *Journal of Education Policy, 34*(4), 470–499. https://doi.org/10.1080/02680939.2018.1460493

Gorur, R. (2016). Seeing like PISA: A cautionary tale about the performativity of international assessments. *European Educational Research Journal, 15*(5), 598–616. https://doi.org/10.1177/1474904116658299

Hastedt, D., & Rocher, T. (2020). *International large-scale assessments in education: A brief guide* (IEA Compass: Briefs in Education No. 10). International Association for the Evaluation of Educational Achievement.

Hayden, J., & McKinlay, D. (2001). Monitoring standards of training. *Medical Education, 35*(1), 68–72.

Ismailov, A. A. (2022). *Tayanch o‘rta ta’lim sifatini baholashning tashkil etish mexanizmini PISA xalqaro dasturi talablari asosida takomillashtirish* [Improving the organisational mechanism for assessing the quality of basic secondary education based on the requirements of the PISA international programme] [Doctoral dissertation, A. Avloniy National Research Institute for Professional Development and Training of Pedagogues in New Methods]. Tashkent.

Kane, M. T., & Wools, S. (2020). Perspectives on the validity of classroom assessments. In S. M. Brookhart & J. H. McMillan (Eds.), *Classroom assessment and educational measurement* (pp. 11–26). Routledge. https://doi.org/10.4324/9780429507533-2

Kenzhetaeva, R., Nurzhanova, S., Beimisheva, A., Stambekova, A., Ryabova, E., & Kulbayeva, D. (2020). Features of the transition from the traditional to the criterial assessment in schools of Kazakhstan. *International Journal of Cognitive Research in Science, Engineering and Education, 8*(3), 73–81. https://doi.org/10.23947/2334-8496-2020-8-3-73-81

Lockheed, M. E., & Wagemaker, H. (2013). International large-scale assessments: Thermometers, whips or useful policy tools? *Research in Comparative and International Education, 8*(3), 296–306. https://doi.org/10.2304/rcie.2013.8.3.296

Lorié, W., & Dadey, N. (2026). *Artificial intelligence in large-scale assessment programs: Applications and considerations for state education agencies*. National Center for the Improvement of Educational Assessment. (ERIC Document ED679214)

Mahmood, Z. (2015). *Comparative effectiveness of internal and external monitoring systems in Punjab education department* [Doctoral dissertation, Hazara University].

Mullis, I. V. S., Martin, M. O., & von Davier, M. (Eds.). (2021). *TIMSS 2023 assessment frameworks*. TIMSS & PIRLS International Study Center, Boston College; International Association for the Evaluation of Educational Achievement.

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

National Academies of Sciences, Engineering, and Medicine. (2019). *Monitoring educational equity*. The National Academies Press. https://doi.org/10.17226/25389

Newton, J. (1999). An evaluation of the impact of external quality monitoring on a higher education college (1993–98). *Assessment & Evaluation in Higher Education, 24*(2), 215–235. https://doi.org/10.1080/0260293990240208

OECD. (2018). *PISA for Development assessment and analytical framework: Reading, mathematics and science*. OECD Publishing. https://doi.org/10.1787/9789264305274-en

OECD. (2023a). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

OECD. (2023b). *PISA 2022 results (Volume II): Learning during – and from – disruption*. OECD Publishing. https://doi.org/10.1787/a97db61c-en

OECD. (2025). *Results from TALIS 2024: The state of teaching*. OECD Publishing. https://doi.org/10.1787/90df6235-en

OECD. (2026). *PISA 2025 results (Volume I): Future-ready students*. OECD Publishing. https://doi.org/10.1787/73451bc5-en

Rubtsova, T. P. (2023). Quality of education: Monitoring and evaluation of educational results. *Vestnik of Samara University. History, Pedagogics, Philology, 29*(4), 66–70. https://doi.org/10.18287/2542-0445-2023-29-4-66-70 (In Russian)

Schleicher, A. (2018). *World class: How to build a 21st-century school system*. OECD Publishing. https://doi.org/10.1787/9789264300002-en

Steiner-Khamsi, G., & Waldow, F. (2018). PISA for scandalisation, PISA for projection: The use of international large-scale assessments in education policy making – an introduction. *Globalisation, Societies and Education, 16*(5), 557–565. https://doi.org/10.1080/14767724.2018.1531234

Turner, D. A., & Li, X. (2026). Catastrophe theory and transitology: A framework for understanding national engagement in international large-scale assessments. *European Education*. Advance online publication. https://doi.org/10.1080/10564934.2026.2627980

von Davier, M., Kennedy, A., Reynolds, K., Fishbein, B., Khorramdel, L., Aldrich, C., Bookbinder, A., Bezirhan, U., & Yin, L. (2024). *TIMSS 2023 international results in mathematics and science*. TIMSS & PIRLS International Study Center, Boston College. https://doi.org/10.6017/lse.tpisc.timss.rs6460

Waldow, F. (2017). Projecting images of the 'good' and the 'bad school': Top scorers in educational large-scale assessments as reference societies. *Compare: A Journal of Comparative and International Education, 47*(5), 647–664. https://doi.org/10.1080/03057925.2016.1262245

Wang, Q. (2026). Construction and empirical analysis of an artificial intelligence-based educational assessment model. *International Journal of Information and Communication Technology Education, 22*(1), 1–16. https://doi.org/10.4018/IJICTE.399169