Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Teacher Qualification Deficits and Mathematics Achievement in Uzbekistan: Multilevel Evidence from PISA 2022

Indira

i.khadjieva@wiut.uz

Westminster International University in Tashkent https://orcid.org/0000-0001-6447-0042

Research Article

Keywords: PISA 2022, hierarchical linear model, teacher qualifications, mathematics achievement, Uzbekistan, rural-urban inequality, transitional economy

Posted Date: July 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10335875/v1

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

Additional Declarations: The authors declare no competing interests.

Abstract

Uzbekistan took part in the OECD Programme for International Student Assessment (PISA) for the first time in 2022, achieving a mean mathematics score of 364 points against an OECD average of 472, with 81% of fifteen-year-old students failing to meet the minimum proficiency benchmark. Despite this result, no prior study has examined the school- and student level determinants of mathematics performance in Uzbekistan using analytically appropriate approaches for hierarchically structured large-scale assessment data. To address this gap, the study draws on the PISA 2022 Uzbekistan dataset (n = 7,293 students; 202 schools) and employs a four-model econometric sequence. The analysis proceeds through: an unconditional two-level HLM for variance decomposition; an OLS baseline with Fay's Balanced Repeated Replication variance estimation; a full two-level HLM with student- and school -level predictors; and simultaneous quantile regression across five quantiles. To account for measurement uncertainty, all models were estimated across ten plausible values and combined following Rubin's (1987) rule for multiply imputed data. Four key findings emerge from the analysis: (a) the intraclass correlation coefficient of 0.181 revealed that a meaningful share of the variation in mathematics performance lies between schools rather than within them, warranting the use of multilevel modelling; (b) at the student level, mathematics self-efficacy, sense of belonging, gender, teacher support, socioeconomic status, and disciplinary climate all showed significant associations with mathematics performance; (c) at the school level, inadequately qualified teaching staff was the only variable that held a consistent and significant negative association with performance across all ten plausible values ($\beta = -10.17$, $p < 0.05$); and (d) a rural subsample analysis revealed that this penalty was 3.4 times larger in rural schools ($\beta = -20.37$) than in urban schools ($\beta = -5.93$), indicating that where students have little access to learning support outside the classroom, the cost of poor teacher quality is far harder to absorb. For Uzbekistan and other transitional education systems, the findings make a case for directing teacher qualification reform where the need – and the cost of inaction- is greatest: rural and undeserved schools.

1. Introduction

Administered triennially by the OECD, PISA has grown into the most widely recognized benchmark for assessing the quality and equity of secondary education systems worldwide (OECD, 2023a). By testing the reasoning and problem-solving skills of fifteen-year-olds in mathematics, reading, and science, PISA offers something national assessments cannot – a shared international framework that allows nations to see where they stand and to identify the school and student factors that shape learning outcomes. Mathematics took center stage in the 2022 cycle, prompting a wave of multilevel studies that sought to understand what – at both the student and school level – drives mathematics performance across OECD and partner countries (Niu et al, 2025; Liu et al, 2024; Frade-Martinez et al, 2024)

Central Asia is largely missing from this literature. Uzbekistan, Kazakhstan, Kyrgyzstan, and Tajikistan share a common Soviet institutional legacy, broadly similar levels of educational access, and a recent record of significant school reform investment – yet they have attracted almost no attention in multilevel PISA research. This is a notable gap, given how analytically interesting these systems are: socioeconomic gradients are comparatively compressed, teacher qualification deficits are acute, and the gap in school quality between rural and urban areas runs deep. Uzbekistan's 2022 PISA debut told a sobering story – a mean mathematics score of 364, 108 points below the OECD average, a 71st place ranking among 80 participating countries and more than four in five students short of the minimum proficiency threshold. Against this backdrop, Uzbekistan has set itself an ambitious target: to rank among the top 30 PISA countries by 2030 – a goal that makes understanding the drivers of current performance not just academically interesting, but practically urgent (Presidential Degree No. DP-21, 2026). Recent reforms – including a 70% salary supplement for certified teachers (Presidential Degree No. DP-79, 2023) and a 46-trillion-soum education budget in 2024 represent a serious commitment of resources. Whether they target the factors that most explain performance gaps between schools, however, is a question that only empirical analysis can answer.

That is what this study sets out to answer. Working with PISA 2022 Uzbekistan dataset – 7,293 students nested within 202 schools – we run four models in sequence: an unconditional two-level HLM, an OLS baseline with Fay's BRR variance estimation, a full two-level HLM, and quantile regression – each chosen to illuminate a different dimension of the student- and school-level determinants of mathematics performance. The analysis follows the OECD's guidelines – estimating all models across ten plausible values, results are pooled using Rubin's (1987) combination rules, and standard errors are computed using BRR variance estimation with 80 replicate weights to reflect the complexity of the sample design.

The study contributes to the literature in three ways. First, it offers the first multilevel analysis of mathematics performance in Uzbekistan using PISA 2022 data, filling gap in the literature on Central Asian education systems. Second, separating the sample by school location shows that students in rural schools bear a disproportionate cost when teachers lack adequate qualifications – a penalty 3.4 times larger than in urban schools (rural $\beta = -20.37$; urban: $\beta = -5.93$) and one that the literature on Uzbekistan has not previously documented. Third, applying quantile regression alongside HLM makes it possible to look inside the distribution rather than just at the average. Teacher support and sense of belonging are most strongly associated with performance at the bottom of the distribution – among the students already falling below the minimum proficiency threshold – while socioeconomic status becomes progressively steeper toward the top of the distribution.

The paper is structured as follows. Section 2 reviews the theoretical framework and empirical literature on the determinants of mathematics achievement in large-scale assessments, with particular attention to multilevel evidence from middle-income and transitional economies. Section 3 describes the PISA 2022 Uzbekistan dataset, sampling design, variable operationalization, and descriptive statistics. Section 4 sets out analytical strategy. Section 5 reports the results across all four models. Section 6 discusses what the findings mean for teacher quality reform in Uzbekistan. Section 7 concludes.

2. Theoretical framework

2.1. The education production function in transitional economies

A long strand of economics of education research, associated most closely with Hanushek (2020), frames learning outcomes as a function of inputs at the student, family, school, and institutional levels – a framework that translates naturally into the multilevel structure of PISA data. What makes this framework particularly relevant is its core empirical claim: that years of schooling tell us surprisingly little about economic outcomes, while measurable cognitive skills tell is a great deal – a finding documented across wide range of countries and decades (Hanushek & Woessmann, 2015). For transitional economies, this distinction is especially consequential. Expanding access to schooling without improving what happens inside classrooms tends to deliver weak human capital gains – and the gap between the return on attendance and the return on actual learning is wider in these settings than almost anywhere else.

Transitional settings, however, present their own complications. Drawing on TIMSS data from seven Eastern European transition countries, Ammermüller, Heijke, and Woessmann (2005) showed that student background dominated performance outcomes, while the contribution of school resources and institutional factors was comparatively modest – a finding that raises questions about how far the standard production function translates across income contexts. Within the sample, a clear divide emerged between countries at different stages of transition. Where market reforms had taken hold and schooling systems had converged toward Western European norms, family background was the stronger predictor. Where transition was still incomplete, school-level factors retained more of their explanatory power – suggesting that the balance between family and school influences shifts as systems mature. Uzbekistan fits neither category. Its reform trajectory has been more gradual than Eastern European transitions Ammermüller et' al. examined, and its institutional context is sufficiently distinct that applying their findings directly – without first testing them against Uzbek data – would risk misreading the system.

2.2. Multilevel analysis of PISA data: methodological considerations

Within any school, students share the same resources, the same leadership, and broadly the same instructional environment – a clustering that has real consequences for how performance differences should be analyzed. This matters methodologically. As Raudenbush and Bryk (2002) show, applying single-level regression to clustered educational data understates standard errors, blurs the distinction between within-school and between-school effects, renders school level influences invalid – a set of problems that multilevel modelling is specifically designed to address.

This is where hierarchical linear modeling comes in. Rather than forcing clustered data into a single-level framework, HLM partitions variance between the student and school levels and estimates effects at each level with uncertainty that honestly reflects the data structure – a combination that has made it the methodological standard in large-scale assessment research.

One of the first questions HLM can answer is how much schools actually matter. The intraclass correlation coefficient addresses this directly, capturing the share of total score variance attributable to difference between schools rather than the differences between students within the same school. Across PISA participants, ICC values range from below 0.10 in Scandinavian countries to around 0.30 on average across OECD members (OECD, 2023b). Middle-income and transitional economies, where school quality is unevenly distributed but student sorting remains incomplete, tend to sit in between. Knowing where Uzbekistan falls is a necessary first step before any school-level analysis can proceed.

The two-level HLM literature on PISA 2022 mathematics is still developing, but early findings point in a consistent direction. Niu, Xu, and YU (2025) examined four high-performing systems – Singapore, Korea, Finland and Denmark – and found that socioeconomic status, mathematics self-efficacy, and teacher support were significant at the student level in each, regardless of how different those systems are in those respects. Frade-Martinez et al (2024) took a three-level approach to Spanish PISA 2022 data, nesting students within schools and schools within autonomous communities. At the student level, socioeconomic background, self-efficacy, and mathematics anxiety were all significant; at the school level, socioeconomic composition exerted an independent effect. Both studies confirm the value of multilevel modelling for PISA data, but neither undertakes beyond high- upper-middle-income contexts – and Central Asia is absent from both. The World Bank (2023) acknowledged Uzbekistan's PISA debut and, despite its last-place ranking among Europe and Central Asia participants, pointed to the significance of having internationally comparable learning outcome data for the country. That data remains, however, without a multilevel econometric treatment in the peer-reviewed literature.

2.3. Teacher qualifications, instructional quality, and student achievement in low- and middle-income settings

What teachers know and how they are trained does not translate into student learning in the same way everywhere. Across country contexts, the relationship between teacher qualifications and outcomes is considerably more varied than the mainstream OECD literature tends to acknowledge. Sanfo (2024) pushes this point further. Using hierarchical linear modelling and PISA-for-Development data from seven low-and middle-income countries, he found that teaching quality did not relate to mathematics achievement in any consistent way – In Ecuador, cognitive support was actually negatively associated with performance, while in the remaining six countries no significant association emerged at all. The contrast with high-income findings is difficult to ignore. This is not simply a matter of effect sizes varying across countries. It reflects something more fundamental: the institutional context in which schools operate shapes how- and whether -instructional factors translate into leaning gains. Findings from OECD systems are a starting point, not a template, for understanding what drives performance in transitional settings like Uzbekistan. PISA does not observe teaching directly. Instead, it relies on students' report of their classroom experience – how much support their teacher provide, how well-ordered lessons are, and how often they are asked to engage with mathematics at a deeper cognitive level. Liu et al. (2024) tracked these three dimensions across the 2012 and 2022 PISA cycles in ten European countries and found a consistent pattern: teacher support, disciplinary climate, and cognitive activation all predicted higher mathematics achievement, and their effects were strongest for students from less advantaged backgrounds, the students, in other words, who have the least to fall back on outside the classroom. The equity implications of this pattern are sharpened considerably in the Uzbekistan context. A system where eight in ten students fall below the minimum proficiency threshold and socioeconomic gradients are compressed is one where instructional quality, if it behaves as it does in Europe, could matter most for the students who are already furthest behind. Testing that assumption against Central Asia data is one of the things this study does.

There is, however, a gap in how the literature operationalizes teacher quality at the school level. Student-reported perceptions of teacher support have been widely studied; principal reports of whether inadequately trained staff are hindering instruction have not. The two capture different things – one reflects the

experience of learning, the other the instructional capacity to deliver it. The proxies available in most multilevel studies – certification rates, years of experience, and student -reported perceptions – tell us something about teacher quality, but they approach it from the outside. What they do not directly measure whether staff qualifications are adequate to the demands of the curriculum. The PISA school questionnaire asks principals a question that goes to the heart of the qualification issue: whether inadequately trained staff are getting in the way of good instruction. This item – SCO17Q02NA- offers a direct school-level measure of qualification deficit, yet it has rarely been treated as a primary predictor in multilevel analyses, least of all for countries participating in PISA for the first time. It is the variable around which this study's central school-level hypothesis is built.

2.4. Rural-urban disparities and the geographic amplification of school quality effects

The rural-urban achievement gap is one of the most consistently documented patterns in comparative education research, but the mechanisms behind it are not uniform. In transitional economies, the drivers tend to differ meaningfully from those identified in high-income settings. In high-income countries, the rural achievement penalty turns out to be largely a story about who lives in rural areas rather than what rural schools do. OECD analysis of PISA and TALIS data found that once socioeconomic status is accounted for, the rural-urban performance gap generally disappears – suggesting that composition, not institutional failure, drives the difference (OECD, 2023b). For policymakers, the distinction matters enormously. Bijou, Liouaeddine, and Bijou (2026), analyzing PISA 2018 and 2022 data across multiple country contexts, show that a compositional gap calls for a different response than an institutional one, and directing resources at the rural school quality, however well intentioned, is unlikely to move the needle if the underlying socioeconomic disadvantage of rural families remains unaddressed.

This picture looks different in transitional economies. Amini and Nivorozhkin (2015) working with PISA data, found that the performance edge of larger urban settlements over rural areas was explained primarily by the socioeconomic composition of their student populations. School resources contributed little or nothing on their own, suggesting that in the transitional contexts, the institutional channel through which urban advantage operates may be weaker than commonly assumed. Compositional factors explain the gap in Russia as well, but Amini and Nivorozhkin (2015) do not leave it there. They point to deep inefficiencies running through the secondary education system, concentrated especially in rural settlements where demographic decline has hollowed out the student population and left schools operating well below capacity. Where Russia's rural school problem is bound up with demographic contraction, Uzbekistan's is rooted in something more tractable, and more urgent. Qualified teachers consistently move toward urban areas, and rural schools are left with what remains: a higher proportion of staff whose training falls short of what the curriculum demands. The gap is structural, labor-market driven, and unlikely to self-correct.

Why does poor teacher quality damage rural students more than urban ones? The compensatory resources hypothesis provides the most grounded answer, and it centers not on what happens inside the classroom, but on what students can draw on outside it. Urban students whose teachers fall short have somewhere else to turn. A tutor can be hired, a parent can step in, a peer group can reinforce what classroom does not, and supplementary services are close enough to use. The school is one node in a wider learning ecosystem, and when it underperforms, the ecosystem has other ways of compensating. In rural Uzbekistan, the ecosystem that cushions urban students against poor teaching simply does not exist in the same form. With fewer tutoring options, weaker home learning environments, and little access to supplementary services, rural students depend on their school, and their teachers in a way that urban students do not. That dependence is precisely what makes qualification deficits so costly where they are most common. Support for this mechanism comes from China, where Sun and Du (2021) found that the leaning gains from teacher training interventions were significantly larger in rural schools than in urban ones. The logic is consistent with the compensatory resources argument: where students have less outside the classroom, what happens inside it carries more weight. This study brings that logic to Uzbekistan for the first time, testing it through a rural-urban subsample comparison that the existing Central Asian literature has left untouched.

Taken together, the absence of multilevel PISA analysis and rural-urban school quality research, not just for Uzbekistan but for all four Central Asian PISA participants, leaves a gap in the comparative education literature that is both obvious and largely unacknowledged. Uzbekistan's 2022 PISA participation gave the field its first internationally comparable window into the country's secondary education system (World Bank, 2023). Thus, study is the first to look through it systematically.

The literature reviewed above leaves three questions unanswered.

RQ1. What proportion of total variance in mathematics performance lies between schools in Uzbekistan, and does the magnitude of between-school variation justify multilevel modelling?

RQ2. Which student-level and school-level variables are significantly associated with mathematics performance after accounting for the clustered structure of the PISA 2022 Uzbekistan data, and how do their effects compare to findings from other PISA participant countries?

RQ3. Does the association between inadequately qualified teaching staff and student mathematics performance vary by school location, and if so, by what magnitude and through what mechanism?

3. Data and Context

3.1. Uzbekistan's secondary education system and PISA 2022 participation

Compulsory schooling in Uzbekistan lasts eleven years. The first nine form the basic general secondary stage; the final two constitute upper secondary education, the level at which PISA assesses fifteen-year-old students. Enrollment in Uzbekistan's Secondary system has reached 6,5 million students across more than 10,500 schools, a near-universal coverage that the system can be justifiably proud of (stat.uz,2024). But enrollment and learning are not the same thing, and three structural constraints stand between the two. First, physical capacity is a binding constraint. Uzbekistan's schools can seat 5.2 million

students, but 6,5 million are enrolled, so 73% of schools operate more than one shift a day, and around a third of all students study in a second or third shift (stat.uz, 2024). The consequence is not just shorter school days, it is reduced access to facilities, heavier teacher loads, and a structural ceiling on the quality of classroom interaction. Second, teacher quality is unevenly distributed across the system, and the distribution follows geography. Certified and experienced teachers concentrate in urban areas, while rural schools disproportionally employ newly qualified or unqualified entrants. With only 24.6% of Uzbekistan's teachers fully certified, the second lowest share among all PISA-participating countries (OECD, 2024), the consequences of this sorting are felt most sharply in the schools least equipped to compensate for it. Third, the growth of private tutoring following economic liberalization has deepened the divide urban and rural students in ways that go beyond what happens inside schools. City students have access to a compensatory ecosystem of tutors, preparatory courses, and supplementary learning centers. For most rural students, that ecosystem does not exist, which means that when their teacher is unqualified, there is nowhere else to turn. This is precisely the mechanism the compensatory resources hypothesis predicts, and it is what the rural subsample analysis in this study sets out to test.

When Uzbekistan's first PISA results arrived in 2022, they confirmed what structural conditions had already suggested, a mean mathematics score of 364 against an OECD average of 472, and a ranking of 71st out of 80 participating nations (OECD, 2023a). Reading and science scores were no more encouraging, 366 and 367 points, respectively. Across all three domains, roughly eight in ten students scored below the OECD Level 2 minimum proficiency threshold, meaning they have not yet acquired the reasoning and literacy skills that further education and the modern labor market take for granted (OECD, 2023a). Uzbekistan placed below all other Central Asian PISA participants, and it is from this starting point that distance to the government's top-30- target by 2030 must be honestly reckoned.

On the policy side, the government has moved with intent. Presidential Degree No. DP-79 (May, 2023) introduced a 70% salary supplement for teachers who earn internationally recognized qualification certificates, an incentive designed to pull more teachers toward formal certification, and in doing so, address the qualification deficit that this study finds to be the single most consistent school-level predictor of lower mathematics performance. Presidential Decree No. DP-21 (2026) translated that ambition into a formal commitment — a top-30 PISA ranking by 2030, backed by a mandate for evidence-based quality monitoring. The logic is sound, but it depends on knowing which factors actually move the needle on student performance. Directing reform at the wrong constraints, however generously funded, will not close a 108-point gap. This study provides the empirical foundation that makes targeted reform possible.

3.2. Dataset and sampling design

The data come from the PISA 2022 international database, which the OECD made publicly available in December 2023. The Uzbekistan subsample was drawn from two source files - the student questionnaire (CY08MSP_STU_QQQ) and the school questionnaire (CY08MSP_SCH_QQQ) — using the country identifier CNT = 'UZB'. The resulting dataset covers 7,293 students across 202 schools, linked at the school level through the international school identifier CNTSCHID.

The sampling process works in two stages. First, students are selected with probability proportional to their size; then, within each selected school, 35 students are drawn. This approach ensures national representativeness while producing the hierarchical data structure, students nested within school, that lies at the heart of the analytical strategy. It also generates intraclass correlation in mathematics scores, which is precisely what the null HLM quantifies. The Uzbekistan sample covers fifteen-year-olds in grade seven and above across urban and rural schools in every region of the country.

3.3. Variables

Dependent variable. Student mathematics performance serves as the outcome variable, measured through ten plausible values (PV1MATH–PV10MATH). The plausible value approach reflects a fundamental feature of PISA's design: no student completes the full item pool, so rather than reporting a single point estimate, PISA draws ten values from the posterior distribution of each student's latent proficiency using item response theory, conditioned on both test responses and background characteristics (OECD, 2024). All models in this study are run separately across each of the ten plausible values and combined using Rubin's (1987) rules, as detailed in Section 4. Mathematics was selected as the primary outcome for two reasons: it is the major domain of PISA 2022, with the deepest item pool and most comprehensive questionnaire coverage, and it is where Uzbekistan's distance from international benchmarks is most pronounced.

Student-level predictors (Level 1). Six variables measured at the student level are drawn from the PISA 2022 student questionnaire. The Index of Economic, Social and Cultural Status (ESCS) captures family socioeconomic background through parental education, occupation, and home educational resources — the broadest available proxy for the material and cultural conditions in which students learn. Gender is coded as a binary variable (1 = male, 0 = female), recoded from the original PISA variable ST004D01T. The remaining four variables are PISA-constructed Weighted Likelihood Estimate indices: mathematics self-efficacy (MATHEFF) measures students' confidence in handling mathematical tasks; disciplinary climate (DISCLIM) reflects classroom order and teacher management of disruption; teacher support (TEACHSUP) captures how attentive, helpful, and instructionally supportive students perceive their mathematics teacher to be; and sense of belonging (BELONG) measures how connected and accepted students feel within their school community.

School-level predictors (Level 2). Five variables from the PISA 2022 school principal questionnaire are included at the school level. The most theoretically central is UNQUALSTAFF, recoded from SC017Q02NA — the item that asks principals how much inadequately or poorly qualified teaching staff hinder instruction. Schools where principals report hindrance to some extent or a lot (categories 3–4) are coded 1; those reporting little or no hindrance (categories 1–2) are coded 0. STAFFSHORT applies the same recoding to SC017Q01NA, which asks about hindrance from a shortage of teaching staff rather than their qualifications. CERTTEACH, drawn from SC018Q02TA01, captures the percentage of full-time teachers who hold full certification; values above 100, which appear in 242 schools and most likely reflect reporting errors, are set to missing. RURAL is recoded from SC001Q01TA, grouping urban and suburban locations (categories 1–2) as 0 and rural and remote locations (categories 3–5) as 1. STRATIO is used as reported in the school file, without modification.

Sampling weights. Because PISA uses a complex stratified cluster design, standard regression weights are insufficient. All models apply the final student weight (W_FSTUWT) to align estimates with the national population, and standard errors are derived using Fay's Balanced Repeated Replication procedure across 80 replicate weights (W_FSTURWT1 through W_FSTURWT80) — the variance estimation approach recommended by the OECD for all PISA-based analyses (OECD, 2024).

3.4. Descriptive statistics

Table 1 summarises the key characteristics of the analytical sample. Uzbekistan's mean mathematics score of 363.67 points (SD = 67.4) sits 108 points below the OECD average of 472 — a gap wide enough to represent roughly one and a half standard deviations of the score distribution. The mean ESCS of −0.68 (SD = 1.01) confirms that Uzbekistan's students fall below the OECD socioeconomic reference point of zero, though the spread of the distribution is notable: students range from well below to moderately above the international mean, reflecting genuine variation in family background within the sample.

At the school level, the data reveal conditions that directly motivate the empirical analysis. More than a quarter of schools (28%) report that instruction is hindered to some extent by inadequately qualified teaching staff, and nearly a quarter (24%) flag staff shortage as an operational constraint. Rural schools account for 36% of the sample — a share large enough to warrant the geographic subsample analysis reported in Section 5.5. The mean student–teacher ratio of 13.79 is relatively low, suggesting that classroom crowding is not the primary resourcing problem in this system. Perhaps the most striking school-level figure is the certified teacher rate: a mean of 11.79% indicates that formal certification — as captured in the PISA school questionnaire — is the exception rather than the rule in Uzbekistan's secondary schools, which helps explain why principal-perceived qualification hindrance, rather than the certification rate itself, emerges as the significant school-level predictor in the models that follow.

Table 1
Descriptive statistics — PISA 2022 Uzbekistan analytical dataset (n = 7,293 students, 202 schools)

Variable	PISA Code	Mean	SD	Min	Max	Expected sign
Dependent variable						
Mathematics score (PV avg.)	PV1–10MATH	363.67	67.4	145.0	625.0	—
Level 1 — Student-level variables						
ESCS index	ESCS	-0.68	1.01	-4.43	2.49	+
Gender (1 = male)	ST004D01T	0.49	0.50	0	1	+ (math)
Math self-efficacy	MATHEFF	-0.42	1.37	-3.73	2.44	+
Disciplinary climate	DISCLIM	0.29	1.12	-2.49	1.85	+
Teacher support	TEACHSUP	0.38	1.19	-2.91	1.56	+
Sense of belonging	BELONG	0.08	1.02	-3.26	2.78	+
Level 2 — School-level variables						
Teaching staff shortage	SC017Q01NA	0.24	0.43	0	1	−
Unqualified staff	SC017Q02NA	0.28	0.45	0	1	−
Certified teachers (%)	SC018Q02TA01	11.79	21.96	0	93	+
Rural location (binary)	SC001Q01TA	0.36	0.48	0	1	−
Student-teacher ratio	STRATIO	13.79	5.74	1.0	28.6	−

Note. PVMATH averaged across PV1MATH–PV10MATH using Rubin (1987) *combination rules. GENDER SD reflects binary coding (male = 1). UNQUALSTAFF and STAFFSHORT are binary indicators recoded from 4-point Likert scale (categories 3–4 = hindrance present). L1 = student level; L2 = school level. Source: PISA 2022 international database (*OECD, 2024*).*

4. Analytical Strategy

4.1. Justification for multilevel modelling

The nested structure of PISA data — students within schools — is not just a design feature; it is an analytical problem for single-level regression. OLS applied to clustered data produces standard errors that are too small, inflates test statistics, and cannot separate what is happening at the student level from what is happening at the school level (Raudenbush & Bryk, 2002). Whether this matters in any given dataset depends on the intraclass correlation coefficient — the proportion of total variance in mathematics scores that lies between schools rather than within them. A low ICC means the clustering can be safely set aside; a high one means it cannot. The null model estimated for Uzbekistan settles the question: the ICC is substantial, and multilevel modelling is necessary rather than optional.

Correct variance estimation requires two things with PISA data: accounting for the complex sampling design and handling the ten plausible values properly. The first is addressed through Fay's Balanced Repeated Replication procedure, implemented with 80 replicate weights and a coefficient of k = 0.5, which produces standard errors that reflect the stratified cluster structure of the sample (OECD, 2024). The BRR variance estimator takes the following form:

$$\varvecV\,(\varvec\theta\,)\;=\;\frac{1}{80\varvecx(0{,}5)^{2}}\sum\nolimits_{\varvecg=1}^{80}(\varvec\theta_{\varvecg}-\varvec\theta\,)^{2}\;=\;1/20\sum\nolimits_{\varvecg=1}^{80}(\varvec\theta_{\varvecg}-\varve$$

$$\varvecS\varvecE\,(\varvec\theta\,)\;=\;\sqrt{\varvecV\,(\varvec\theta\,)}$$

where θ is the full-sample estimate, $\varvec\theta_{\varvecg}$ is the estimate from the g-th replicate sample (g = 1, ..., G = 80), k = 0.5 is the Fay adjustment factor, and G = 80 is the number of replicates.

The second is handled by estimating each model ten times — once per plausible value — and combining the results using Rubin's (1987) rules, which separate the uncertainty that comes from sampling from the uncertainty that comes from the plausible value imputation process.

$$\bar{\beta}=\;\frac{1}{M}\sum\nolimits_{m=1}^{M}\beta_{m}$$

$$T=\bar{U}+\left(1+\frac{1}{M}\right)B$$

where β_{m} denotes the coefficient estimated using plausible value m, $\bar{U}$ is the average within-imputation variance, and B is the between-imputation variance. Taking a shortcut, using a single plausible value or averaging across ten, would understate the true variance and inflate significance; the Rubin (1987) procedure ensures this does not happen.

4.2. Model sequence

The analysis proceeds through four models, each building on what the previous one establishes. The null HLM starts with no predictors and is estimated by restricted maximum likelihood; its sole purpose is to decompose total mathematics score variance into within-school and between-school components, yielding the ICC that addresses RQ1. The outcomes of this preliminary model provided insights into the distribution of variances across different levels, serving as a foundational assessment before introducing predictor variables.

$$Y_{ij}=\gamma_{00}+\epsilon_{ij}+r_{0j}$$

Where

Y_{ij} is the mathematics plausible value for the student i in school j

γ_{00} is the grand mean mathematics score s

ϵ_{ij} is the residuals for student i in school j

r_{0jk} is the school level residual

The key diagnostic statistic derived from Model 1 is the Intraclass Correlation Coefficient (ICC), which expresses the proportion of total score variance attributable to differences between schools:

$$\varvecI\varvecC\varvecC_{\varvecs\varvecc\varvech\varveco\varveco\varvecl}=\frac{\varvec\tau^{2}}{\varvec\tau^{2}\;\;+\varvec\sigma^{2}}$$

where τ² is the between-school variance and σ² is the within-school variance. A high ICC signals that school identity is an important determinant of individual student outcomes — and thus that multilevel modelling is not merely appropriate but necessary.

The OLS baseline then regresses scores on the six student-level predictors with BRR variance estimation — not because OLS is the preferred method, but because it offers a transparent benchmark for cross-study comparison and because contrasting its ESCS coefficient with the full HLM estimate reveals how much of the socioeconomic gradient is compositional. The estimated equation is:

$$Y_{ij}=\beta_{0j}+\beta_{1}ESCS_{ij}+\beta_{2}Gender_{ij}+\beta_{3}MATHEFF_{ij}+\beta_{4}DISCLIM_{ij}+\;\beta_{5}TEACHSUP_{ij}\;+\beta_{6}BELONG_{ij}\;+\;\epsilon_{ij}$$

where Y_{ij} is the (averaged) plausible value mathematics score for student i in school j; α is the intercept; β_1 through β_6 are partial regression coefficients representing the expected change in mathematics score associated with a one-unit increase in the respective predictor, holding all other predictors constant; and ϵ_{ij} is the error term.

The two-level HLM is the study's primary model: student-level predictors enter at Level 1, school-level predictors at Level 2, and the model is estimated by maximum likelihood. It addresses RQ2 directly and is the source of the findings reported in Section 5.

The full two-level HLM is specified as follows:

$$Level\ 1$$

$$Y_{ij} = \beta_{0j} + \beta_1 ESCS_{ij} + \beta_2 Gender_{ij} + \beta_3 MATHEFF_{ij} + \beta_4 DISCLIM_{ij} + \beta_5 TEACHSUP_{ij} + \beta_6 BELONG_{ij} + \varepsilon_{ij}$$

$$Level\ 2$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} STAFFSHORT_j + \gamma_{02} UNQUALSTAFF_j + \gamma_{03} CERTTEACH_j + \gamma_{04} RURAL_j + \gamma_{05} STRATIO_j + u_{0j}$$

Combining these two levels yields the reduced-form mixed model equation:

$$Y_{ij} = \gamma_{00} + \gamma_{01} STAFFSHORT_j + \gamma_{02} UNQUALSTAFF_j + \gamma_{03} CERTTEACH_j + \gamma_{04} RURAL_j + \gamma_{05} STRATIO_j + \beta_1 ESCS_{ij} + \beta_2 Gender_{ij} + \beta_3 MATHEFF_{ij} + \beta_4 DISCLIM_{ij} + \beta_5 TEACHSUP_{ij} + \beta_6 BELONG_{ij} + u_{0j} + \varepsilon_{ij}$$

All student-level continuous predictors (ESCS, MATHEFF, DISCLIM, TEACHSUP, BELONG) are group-mean centred — that is, centred on the school mean — to separate within-school from between-school effects and to reduce multicollinearity between student-level and school-level terms. School-level continuous variables (CERTTEACH, STRATIO) are grand-mean centred.

The final model — simultaneous quantile regression at five quantiles (τ = 0.10, 0.25, 0.50, 0.75, 0.90) with 100 bootstrap replications — asks whether the relationships identified in the HLM look the same across the full performance distribution, or whether they are stronger or weaker among the students currently scoring at the bottom, where eight in ten of Uzbekistan's fifteen-year-olds are concentrated. Quantile regression, introduced by Koenker and Bassett (1978) and applied to education data by Koenker and Hallock (2001), estimates the conditional quantile function $Q_Y(\tau \mid \mathrm{X}) = X^T \beta(\tau)\ at$ selected quantiles τ across the performance distribution. This approach addresses the limitation of OLS and HLM mean estimates, which may conceal important variation across the distribution. Policymakers using only mean estimates risk choosing 'one size fits all' interventions that miss the distributional structure of the problem. This concern is acute for Uzbekistan, where 81% of students already score below the minimum proficiency threshold.

Two further checks are run after the primary models. Re-estimating the full HLM without mathematics self-efficacy tests whether the variable is mediating — rather than independently predicting — mathematics performance, which would affect how the student-level coefficients should be interpreted. Estimating separate HLMs for urban and rural subsamples then provides the direct evidence needed to answer RQ3: whether the association between inadequately qualified teaching staff and student performance is larger in rural schools, and by how much.

All estimation is carried out in Stata 17. Hierarchical linear models are fitted using the mixed command and quantile regression using sqreg. Before any models were estimated, the raw PISA 2022 files required preparation: SAS source files were converted to Stata format, student and school records were merged using the CNTSCHID identifier, and UNQUALSTAFF, STAFFSHORT, and RURAL were recoded as binary variables. The PISA 2022 international database is publicly available from the OECD data portal

5. Results

5.1. Null model and variance decomposition

The null model establishes the answer to RQ1. Between-school variance is 810.44 and within-school variance is 3,672.62, giving an ICC of 0.181. This means that nearly a fifth of all variation in Uzbekistan's mathematics scores is accounted for by which school a student happens to attend, a finding that sits above the intermediate range typical of middle-income and transitional economies and confirms that school-level factors are a meaningful source of performance inequality in this system. The grand mean of 363.7 points anchors the distributional analysis that follows.

The likelihood ratio test leaves no ambiguity: across all ten plausible values, chibar² ranges from 936 to 1,052 and the p-value is below 0.001 in every case. The between-school variance is real, and multilevel modelling is not a methodological preference here — it is a requirement. The ICC of 0.181 puts Uzbekistan in an interesting comparative position: higher than Scandinavian systems, where schools are so uniform that attending one rather than another barely matters, but considerably lower than highly tracked systems such as Germany or Hungary, where ICCs regularly exceed 0.50. It is the signature of a system where school quality varies substantially but where students have not yet been sorted sharply across schools by background or ability.

Table 2
Null model variance decomposition — Uzbekistan PISA 2022 mathematics (averaged across 10 PVs)

Component	Estimate	Share of total variance
Between-school variance (τ^2)	810.44	18.1%
Within-school variance (σ^2)	3,672.62	81.9%
Total variance	4,483.06	100%
Intraclass Correlation Coefficient (ICC)	0.181	—
Grand mean mathematics score	363.67 points	—
LR test vs. linear model (range)	chibar2: 936–1,052	$p < 0.0001$ across all 10 PVs

Note. All variance components are estimated by restricted maximum likelihood using the mixed command in Stata 17 and averaged across the ten plausible mathematics values (PV1MATH–PV10MATH) following Rubin's (1987) combination rules. The likelihood ratio test assesses the improvement in fit of the two-level model over a single-level linear regression.

5.2. OLS baseline — student-level predictors

The OLS baseline tells a consistent story across all six student-level predictors — each is positively and significantly associated with mathematics performance (n = 5,007, 202 clusters, average $R^2 = 0.096$). Mathematics self-efficacy stands out as the strongest individual predictor: each additional unit on the MATHEFF index is associated with roughly 8.3 more score points ($\beta = 8.30$, $p < 0.001$). In a system where the average student scores 108 points below the OECD mean, that is not a trivial margin. Sense of belonging is the second strongest student-level predictor ($\beta = 6.77$, $p < 0.001$), and its prominence at the baseline stage — before school-level variables have been introduced — is telling. It suggests that the social and emotional dimensions of school experience are not peripheral to mathematics learning in Uzbekistan; they are among its most important student-level correlates. The gender coefficient ($\beta = 11.05$, $p < 0.001$) indicates an 11-point male advantage in mathematics, consistent with the pattern observed across most PISA participants at age fifteen, though the size of the gap in Uzbekistan warrants attention in its own right. Socioeconomic status is a significant predictor ($\beta = 8.36$, $p < 0.001$), but the size of the coefficient understates how unusual Uzbekistan's socioeconomic gradient actually is. ESCS explains only around 2% of total score variance — a fraction of the 15% OECD average. In most PISA-participating countries, family background is the single most powerful predictor of student performance; in Uzbekistan, it barely registers as a source of variation. This compression has a clear analytical consequence: if family background cannot account for the between-school differences the ICC reveals, institutional factors at the school level must. That is precisely what the primary HLM model sets out to establish.

Table 3
OLS baseline results — averaged across ten plausible values (n = 5,007 students, 202 schools)

Variable	Coefficient (avg.)	p-value	Interpretation
ESCS	8.36	< 0.001	1-unit SES increase → 8.4 score points
GENDER (male = 1)	11.05	< 0.001	Male students score 11 points higher in math
MATHEFF	8.30	< 0.001	Strongest student-level predictor
TEACHSUP	5.46	< 0.001	Teacher support → 5.5 points higher
BELONG	6.77	< 0.001	School belonging → 6.8 points higher
DISCLIM	2.64	< 0.05	Classroom order → 2.6 points higher
Constant	363.71	< 0.001	—
R^2	0.096	—	9.6% of total variance explained

Note. All standard errors are estimated using Fay's Balanced Repeated Replication procedure with 80 replicate weights, clustered at the school level across 202 schools. Coefficients and standard errors are averaged across the ten plausible mathematics values (PV1MATH–PV10MATH) following Rubin's (1987) combination rules. The disciplinary climate index (DISCLIM) is significant at the conventional $p < 0.05$ threshold in eight of the ten plausible values and marginally significant ($p < 0.10$) in the remaining two; it is retained in the model on theoretical grounds.

5.3. Two-level HLM — primary model

The full two-level HLM (n = 4,653 students, 190 schools) brings the student- and school-level picture together. At the student level, all six predictors hold their ground — same signs, same significance — after the school context is accounted for. Mathematics self-efficacy remains the dominant influence ($\beta = 7.37$, $p < 0.001$), with belonging ($\beta = 6.30$, $p < 0.001$), gender ($\beta = 9.12$, $p < 0.001$), teacher support ($\beta = 4.63$, $p < 0.001$), ESCS ($\beta = 3.87$, $p < 0.001$), and disciplinary climate ($\beta = 1.70$, marginal) following in order. The most analytically significant shift is in the ESCS coefficient, which falls by 54% — from 8.36 in the OLS baseline to 3.87 in the HLM. Nearly half of the socioeconomic gradient visible in single-level analysis is not a within-school effect at all; it is a between-school sorting effect, with disadvantaged students disproportionately concentrated in lower-performing schools. That finding alone justifies the multilevel approach — and it sharpens the question of what is actually driving differences between schools.

The Level 2 results answer RQ2 with unusual clarity. Of the five school-level predictors, only UNQUALSTAFF — the binary indicator of principal-reported hindrance from inadequately qualified teaching staff — achieves statistical significance across all ten plausible values ($\beta = -10.17$, $p < 0.05$). The effect is substantively meaningful: schools where teacher qualifications are reported as a hindrance score around 10 points lower in mathematics after controlling for student-level composition, equivalent to approximately 15% of a standard deviation. Staff shortage, teacher certification rate, student-teacher ratio, and

school location each fall short of consistent significance. The implication is direct: across all the institutional inputs available in the PISA 2022 school questionnaire, it is not resources, not school size, and not location that most reliably predicts where a school sits in the performance distribution — it is whether its teachers are adequately qualified to teach.

The other school-level variables do not reach significance. Staff shortage (β = 6.98, p > 0.10), teacher certification rate (β = −0.09, p > 0.10), and student-teacher ratio (β = 0.09, p > 0.10) all fall short — and their non-significance is itself informative. It is not the raw supply of teachers, nor the aggregate certification rate, that predicts school performance; it is whether the teachers who are present are adequately qualified for what they are being asked to teach. School location tells a more complex story. Rural schools score around 12 points higher than urban schools after student composition is controlled (β = 12.08, p < 0.01) — a finding that runs against intuition but becomes more interpretable when the structure of the urban school distribution is considered. Urban schools in Uzbekistan are large and varied, and the weakest among them are numerous enough to pull the urban mean below the rural average, even though rural schools face greater qualification deficits on average. The inclusion of school-level predictors reduces between-school variance by 27.9%, from 810.44 in the null model to 584.05 in the full HLM — a meaningful reduction that confirms the Level 2 specification is doing real explanatory work.

Table 4
Two-level HLM results — averaged across ten plausible values (n = 4,653 students, 190 schools)

Variable	Coefficient (avg.)	p-value	Interpretation
Level 1 — Student-level fixed effects			
ESCS	3.87	< 0.001	Within-school SES gradient (halved from OLS 8.36)
GENDER (male = 1)	9.12	< 0.001	Male advantage in mathematics
MATHEFF	7.37	< 0.001	Strongest student-level predictor
TEACHSUP	4.63	< 0.001	Teacher support — significant and positive
BELONG	6.30	< 0.001	School belonging — significant and positive
DISCLIM	1.70	Marginal	Classroom order — marginally significant
Level 2 — School-level fixed effects			
STAFFSHORT	6.98	p > 0.10	Staff quantity shortage — not significant
UNQUALSTAFF	-10.17	< 0.05	Unqualified staff — significant and negative ✓
CERTTEACH	-0.09	p > 0.10	Teacher certification rate — not significant
RURAL	12.08	< 0.01	Rural schools score higher — see text
STRATIO	0.09	p > 0.10	Student-teacher ratio — not significant
Variance components			
Between-school variance (τ^2)	584.05	—	Reduced by 27.9% from null model
Within-school variance (σ^2)	3,369.98	—	Reduced by 8.2% from null model
Conditional ICC	0.147	—	14.7% residual between-school variance

Note. Models are estimated by maximum likelihood using the mixed command in Stata 17. All coefficients and standard errors are averaged across the ten plausible mathematics values (PV1MATH–PV10MATH) following Rubin's (1987) combination rules. UNQUALSTAFF is negative and statistically significant at p < 0.05 in every plausible value, with coefficients ranging from − 8.48 to − 11.23, confirming the consistency of the teacher qualification effect across the full set of draws. RURAL is positive and significant across all ten plausible values, with coefficients ranging from + 10.37 to + 13.60. Variables described as marginal achieve p < 0.10 in seven of the ten plausible values.

5.4. Quantile regression — distributional effects

The quantile regression results complicate the average-based picture in ways that matter for policy. Associations between key predictors and mathematics performance are not stable across the score distribution — they shift, sometimes substantially, between the bottom and the top. Teacher support illustrates this most clearly. At Q10, where students are deepest below the minimum proficiency threshold, each unit of teacher support is associated with nearly 6 additional score points (β = 5.97, p < 0.001). That association weakens steadily as performance rises — 3.72 at Q75 — and by Q90 it has fallen to 2.06 and lost statistical significance entirely. The students for whom teacher support matters most are precisely those who are already struggling most.

The pattern for belonging mirrors that of teacher support: β = 8.49 at Q10, falling to β = 4.97 at Q90 — further evidence that school climate effects are concentrated among weaker performers. ESCS moves in the opposite and more troubling direction: its association with mathematics performance rises from β = 4.76 at Q10 to β = 11.02 at Q90, meaning that socioeconomic advantage exerts its strongest influence not among the most disadvantaged students but among those already performing well. The distribution is being pulled apart from both ends. UNQUALSTAFF, by contrast, is consistent across the full distribution — negative and significant at all five quantiles (range: −11.66 to − 13.96) — confirming that poorly qualified teachers impose a performance cost regardless of where in the distribution a student sits. For policy, the distributional results carry a message that aggregate HLM estimates cannot convey: interventions targeting teacher support and school belonging generate their largest returns precisely among the students Uzbekistan can least afford to leave behind — those in the 81% currently below the minimum proficiency threshold.

Table 5
Quantile regression results — five quantiles (PV1MATH, n = 4,653 students)

Variable	Q10	Q25	Q50	Q75	Q90
ESCS	4.76***	4.48***	7.86***	9.68***	11.02***
GENDER	1.79	5.44**	8.54***	12.91***	17.34***
MATHEFF	3.35***	4.39***	6.56***	9.38***	12.51***
DISCLIM	2.85**	4.04***	3.77***	2.59**	1.38
TEACHSUP	5.97***	5.00***	6.56***	3.72***	2.06
BELONG	8.49***	7.56***	5.35***	7.42***	4.97**
STAFFSHORT	10.04**	5.52	5.58*	12.19***	20.32***
UNQUALSTAFF	-11.66***	-11.52***	-7.62**	-10.70***	-13.96***
CERTTEACH	-0.11	-0.06	-0.07	-0.10*	-0.16*
RURAL	8.40***	11.25***	9.86***	12.58***	9.94***
STRATIO	-0.10	0.05	0.01	0.12	0.12
Pseudo R^2	0.041	0.040	0.048	0.062	0.082

Note. *** p < 0.001, ** p < 0.05, * p < 0.10. Estimated using sqreg with 100 bootstrap replications in Stata 17. Outcome: PV1MATH. Standard errors from 100 bootstrap replications.

5.5. Robustness checks — urban and rural subsamples

The rural-urban subsample analysis is where RQ3 is answered most directly. Estimating separate HLMs for urban schools (RURAL = 0, n = 3,019, 132 schools) and rural schools (RURAL = 1, n = 1,634, 58 schools) produces a geographic contrast that the pooled model averages away. In urban schools, the UNQUALSTAFF coefficient is − 5.93 on average across ten plausible values, and it does not reach conventional significance in most specifications. A negative association is present, but its inconsistency across plausible values suggests that the urban school context — with its wider array of compensatory resources — moderates the performance cost of inadequate teacher qualifications in ways that rural schools cannot replicate.

Rural schools tell a very different story. The UNQUALSTAFF coefficient averages − 20.37 — significant in seven of ten plausible values ($p < 0.05$) — and is 3.4 times larger than the equivalent urban coefficient. The gap is consistent with what the compensatory resources hypothesis predicts. Urban students facing poorly qualified teachers have options: private tutors, better-equipped homes, peers who reinforce academic work outside school hours. For most rural students in Uzbekistan, those options do not exist. The school — and the teacher within it — is the primary, often the only, source of structured learning. When that teacher is inadequately qualified, the performance cost has nowhere to be absorbed. The answer to RQ3 could hardly be plainer: the teacher qualification penalty is not a uniform school-level effect but a geographically concentrated one, and Presidential Decree No. DP-79's salary supplement for certified teachers will deliver its greatest returns precisely in the rural schools where the deficit runs deepest.

Table 6
Robustness check — UNQUALSTAFF coefficient: full sample, urban, and rural subsamples (averaged across ten plausible values)

Subsample	n (students)	Schools	UNQUALSTAFF β	Significant?
Full sample (primary model)	4,653	190	−10.17	Yes — p < 0.05 in 8/10 PVs
Urban (RURAL = 0)	3,019	132	−5.93	No — p < 0.05 in 3/10 PVs
Rural (RURAL = 1)	1,634	58	−20.37	Yes — p < 0.05 in 7/10 PVs

Note. All estimates averaged across PV1MATH–PV10MATH using Rubin (1987) rules. RURAL variable excluded from subsample models. Rural penalty (− 20.37) is 3.4 times larger than urban penalty (− 5.93), consistent with the compensatory resources hypothesis.

6. Discussion

6.1. What the ICC tells us about school quality variation in Uzbekistan

The null model ICC of 0.181 does more than justify the multilevel approach — it tells us something important about how Uzbekistan's school system works. Nearly one in five points of variation in mathematics performance is explained not by anything a student brings to school but by which school they happen to attend. Translated into score terms, that is roughly a 17-point difference between a student at an average school and one at a school a standard deviation above the mean on the between-school distribution — a gap that is modest by the standards of highly tracked systems but far from trivial in a context where the national mean already falls 108 points short of the OECD average.

The comparative picture clarifies what the ICC of 0.181 actually represents. It is higher than Scandinavian systems — where school uniformity makes institutional differences almost irrelevant — but far below the tracked systems of Germany or Hungary, where early ability grouping concentrates students so sharply that ICCs routinely exceed 0.50 (OECD, 2023b). Uzbekistan occupies the space in between: a system where school quality varies enough to matter

but where the sorting of students across schools by socioeconomic background remains limited. The 2% of variance explained by ESCS is the clearest evidence of this — students from different family backgrounds are, by and large, attending the same kinds of schools. For policymakers, this is an important structural feature. It means that the gap between Uzbekistan's best- and worst-performing schools is not primarily a story about who attends them; it is a story about what happens inside them. The Level 2 results are where that story is told.

6.2. The UNQUALSTAFF finding in comparative perspective

Among all the school-level factors available in the PISA 2022 questionnaire, one stands apart. Inadequately qualified teaching staff is the only variable that consistently and significantly predicts mathematics performance across all ten plausible values — and the consistency of that finding, not just its magnitude, is what makes it the central empirical contribution of this study. Schools where principals report that poorly qualified teachers hinder instruction score around 10 points lower in mathematics after student composition is accounted for ($\beta = -10.17$), equivalent to roughly 15% of a standard deviation or about one-third of the per-standard-deviation gap between Uzbekistan and the OECD average. Staff shortage does not reach significance. Neither does the student-teacher ratio, nor the formal certification rate. What predicts school performance in Uzbekistan is not how many teachers there are, nor whether they hold a certificate, but whether the teachers standing in classrooms are adequately qualified for what they are being asked to teach.

The international evidence places the Uzbekistan finding in company. Pholphirul et al. (2023) found that teacher shortages significantly damage performance in Thailand, with rural schools bearing the heaviest cost — a pattern that tracks the Uzbekistan results closely and suggests that teacher quality deficits follow similar pathways across middle-income Asian education systems. More broadly, the OECD (2023b) found that principal-reported instructional hindrance from staffing deficits increased across 58 countries between 2018 and 2022 — a post-pandemic deterioration that makes Uzbekistan's challenge representative rather than exceptional. What the Uzbekistan data add to this picture is precision. General staff shortage — STAFFSHORT — is non-significant in the HLM. It is qualification adequacy specifically — UNQUALSTAFF — that predicts school performance. The implication for reform is unusually clear: the problem is not that Uzbekistan's classrooms are understaffed; it is that too many of the teachers in them are inadequately prepared for what they are being asked to teach. Expanding the teacher workforce without raising its qualification level will not close the gap. Presidential Decree No. DP-79 (2023) — which makes certification financially worthwhile through a 70% salary supplement — is the kind of targeted intervention the evidence actually supports.

The non-significance of the formal certification rate is not a null finding to be set aside — it is a finding in its own right. A certificate issued by a domestic authority tells us that a teacher has cleared an administrative bar; it does not tell us whether they can teach mathematics to a fifteen-year-old. In Uzbekistan, where certification has long served as an entry credential rather than a meaningful signal of classroom readiness, the two may have drifted far apart. The government appears to have sensed this: the decision to tie Presidential Decree No. DP-79's salary supplement to internationally recognised certificates rather than domestic ones is, in effect, an acknowledgement that the existing credentialing system does not reliably capture what it is supposed to measure. UNQUALSTAFF cuts through that problem by asking principals directly whether qualification levels are hindering instruction — a more immediate and operationally grounded measure of the quality constraint. The data confirm it is the right question to ask.

6.3. The rural amplification — implications for geographic targeting

The rural subsample analysis is where the study's second major contribution comes into focus. In rural schools, the performance penalty associated with inadequately qualified teaching staff is − 20.37 — 3.4 times the urban estimate of − 5.93. That is not a number that can be dismissed as statistical noise: it holds across seven of ten plausible values in the rural subsample, and it fits the pattern that research on teacher quality and school location in middle-income countries has been building toward.

The compensatory resources hypothesis makes sense of the gap. In Uzbekistan's cities, economic liberalisation has fuelled a substantial private tutoring market — particularly in Tashkent — and urban students whose classroom instruction falls short can draw on tutors, better-resourced homes, and peer networks that reinforce academic work outside school hours. For most rural students, those options simply do not exist. The school is where learning happens, and the teacher is the primary vehicle through which it does or does not. When that teacher is inadequately qualified, there is nothing else in the rural environment to compensate — and the performance penalty is correspondingly larger. Pholphirul et al. (2023) found a closely parallel pattern in Thailand, where teacher shortages carry a particularly severe cost for rural students. Taken together, the Thai and Uzbek evidence suggests that the rural amplification of teacher quality penalties is not a local anomaly but a structural regularity in middle-income education systems where geographic inequalities in compensatory resources run deep.

The geography of the finding has direct consequences for reform design. Uzbekistan's rural schools account for 44% of all schools in the system — and they are, on average, both more likely to employ inadequately qualified teachers and more severely affected when they do. The certification incentive under Presidential Decree No. DP-79 needs to reach these schools in meaningful numbers. The danger — well documented in middle-income country reform experience — is that salary supplements tied to certification tend to flow toward urban teachers first. They are better placed to find out about the process, better supported in preparing for it, and more likely to face competitive pressure that makes the incentive worth pursuing. If that pattern repeats itself in Uzbekistan, the rural amplification this study documents will not diminish — it will deepen. Tracking rural certification uptake as an annual indicator, as the monitoring framework in Chapter III proposes, is a practical response to that risk. For middle-income systems more broadly, the implication is hard to avoid: in a school system with pronounced rural-urban disparities, teacher quality reform that is not geographically targeted is teacher quality reform that will miss the students who need it most.

6.4. Limitations

Three limitations are worth naming directly. The first is one the data impose rather than choices the analyst makes: the analysis is cross-sectional. Because 2022 was Uzbekistan's first PISA cycle, there is no prior wave to compare against, and the findings are confined to conditional associations rather than causal estimates. The coefficients describe a pattern across schools — schools with inadequately qualified staff score around 10 points lower than comparable schools without — but they cannot establish that improving teacher qualifications would produce a 10-point gain. That question requires

variation over time, not just across schools. Uzbekistan's second PISA cycle, expected in 2025, will provide that variation and create the conditions for a difference-in-differences evaluation of whether the teacher certification incentives under Presidential Decree No. DP-79 have moved performance in the schools and regions where uptake was greatest.

The second limitation concerns how rurality is measured. A binary variable derived from a single principal questionnaire item captures the broad distinction between urban and rural schools, but rural schooling in Uzbekistan is far from uniform. A school in a peri-urban district centre is rural in the same coding as a school in a remote mountain village — yet the conditions they face, and the degree to which they can compensate for teacher qualification deficits, are likely very different. Multi-shift operation rates, distance from the nearest city, and how tight the local teacher labour market is are all plausible moderators of the amplification effect, and none is captured by a binary RURAL indicator. Extending the analysis to a three-level model — students within schools within regions — would allow regional variation in the rural penalty to be estimated directly and would give policymakers a more granular picture of where the certification reform needs to reach most urgently.

The third limitation is one shared by all school questionnaire-based research: the measures are what principals report, not what an independent observer would record. UNQUALSTAFF tells us that a principal believes inadequately qualified staff are hindering instruction — it does not tell us what a formal audit of teaching qualifications would find, or whether two principals facing similar staffing profiles would describe them in the same way. Principal perceptions are a well-validated proxy for school conditions in the international large-scale assessment literature, and there is no reason to think they are systematically misleading here. But they are perceptions, and the gap between perception and objective measure is a source of noise that better data would reduce. Linking PISA school questionnaire records with Uzbekistan's national administrative data on teacher qualifications and inspection outcomes — something national research institutions are well-placed to do — would produce a more precise measure of the qualification constraint. Paired with the longitudinal variation that PISA 2025 will bring, that linkage would create the foundation for instrumental variable estimation and, ultimately, for the kind of causal identification that this study's cross-sectional design cannot yet provide.

7. Conclusion

Applying a four-model econometric sequence to the PISA 2022 Uzbekistan dataset, this study identifies three principal findings. An ICC of 0.181 confirms that school attended accounts for 18.1% of total mathematics score variance, establishing the analytical necessity of multilevel modelling. The two-level HLM finds that inadequately qualified teaching staff is the only school-level variable consistently and significantly associated with lower mathematics performance across all ten plausible values ($\beta = -10.17$, $p < 0.05$), with staff shortage, certification rate, student-teacher ratio, and school location all non-significant. The rural subsample analysis shows that the teacher qualification penalty is 3.4 times larger in rural schools ($\beta = -20.37$) than urban ones ($\beta = -5.93$), consistent with the compensatory resources hypothesis.

The findings point clearly to where Uzbekistan's reform effort should be directed. Presidential Decree No. DP-79 (May 2023) — the 70% salary supplement for internationally certified teachers — is the right instrument: the evidence identifies teacher qualification adequacy, not staff numbers or domestic credentials, as the binding school-level constraint. The non-significance of CERTTEACH alongside the significance of UNQUALSTAFF suggests that existing domestic certification frameworks do not reliably signal the pedagogical competencies that actually drive performance differences between schools. Presidential Decree No. DP-21 (2026) provides the legislative context within which these findings should shape resource allocation decisions. The 3.4 times rural amplification documented here makes the geographic implication concrete: the certification pathway needs to reach rural schools first and most intensively, because that is where improving teacher quality will deliver its largest return.

The study's limitations point to two concrete research directions. The more pressing is longitudinal. Uzbekistan's PISA 2025 cycle, due in late 2026, will provide the first real test of whether Presidential Decree No. DP-79's certification incentive has moved the needle — reducing the share of schools where principal-reported qualification deficits hinder instruction, and producing score gains of a magnitude consistent with the HLM estimates. A difference-in-differences design exploiting variation in certification uptake across schools and regions would, for the first time, support causal rather than associational claims about the relationship between teacher qualification adequacy and student performance in Uzbekistan. The second direction is geographic. A three-level extension of the current framework — students within schools within regions — would decompose the between-school variance further, separating what is attributable to school-level conditions from what reflects the administrative and economic context of the region in which a school sits. That decomposition would not only advance the analysis but give policymakers a more precise answer to the question this study raises but cannot fully resolve: not just which schools face the largest teacher qualification deficit, but where in Uzbekistan's regions those schools are concentrated.

Declarations

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this paper, the author used Claude to assist with translation from UZBEK and Russian sources and to support the editing and prose refinement of the manuscript. The author carefully reviewed and edited all content produced with this tool to maintain academic rigor and takes responsibility for the content of published article.

References

1. Ammermüller A, Heijke H, Woessmann L (2005) Schooling quality in Eastern Europe: Educational production during transition. Econ Educ Rev 24(5):579–599. https://doi.org/10.1016/j.econedurev.2004.08.012

2. Amini C, Nivorozhkin E (2015) The urban–rural divide in educational outcomes: Evidence from Russia. Int J Educational Dev 44:118–133. https://doi.org/10.1016/j.ijedudev.2015.07.006
3. Bijou N, Liouaeddine M, Bijou M (2026) Unequal Starts: How Socioeconomic Status Shapes Child and Adolescent Learning Outcomes–International Insights from PISA 2018 and 2022. Child Indic Res, 1–47
4. Frade-Martínez C, O'Hara J, Gamazo A, Olmos-Migueláñez S, Brown M (2024) A multilevel investigation of factors related to achievement in Ireland and Spain using PISA data. Front Educ 9., Article 1306197. https://doi.org/10.3389/feduc.2024.1306197
5. Hanushek EA (1979) Conceptual and empirical issues in the estimation of educational production functions. J Hum Resour 14(3):351–388. https://doi.org/10.2307/145575
6. Hanushek EA (2003) The failure of input-based schooling policies. Econ J 113(485):F64–F98. https://doi.org/10.1111/1468-0297.00099
7. Hanushek EA (2020) Education production functions. In S. Bradley & C. Green (Eds.), *The economics of education: A comprehensive overview* (2nd ed., pp. 161–170). Academic Press. https://doi.org/10.1016/B978-0-12-815391-8.00013-6
8. Hanushek EA, Woessmann L (2015) The knowledge capital of nations: Education and the economics of growth. MIT Press
9. Koenker R, Bassett G (1978) Regression quantiles. Econometrica 46(1):33–50. https://doi.org/10.2307/1913643
10. Koenker R, Hallock KF (2001) Quantile regression. J economic Perspect 15(4):143–156
11. Liu X, Yang Hansen K, Valcke M, De Neve J (2024) A decade of PISA: Student perceived instructional quality and mathematics achievement across European countries. ZDM – Math Educ 56:859–891. https://doi.org/10.1007/s11858-024-01630-7
12. Niu J, Xu H, Yu J (2025) Identifying multilevel factors on student mathematics performance for Singapore, Korea, Finland, and Denmark in PISA 2022: Considering individualistic versus collectivistic cultures. Humanit Social Sci Commun 12:151. https://doi.org/10.1057/s41599-025-04466-y
13. OECD (2019) Learning in rural schools: Insights from PISA, TALIS and the literature. OECD Publishing. https://doi.org/10.1787/8b1a5cb9-en
14. OECD (2023a) PISA 2022 results (I): The state of learning and equity in education. OECD Publishing. https://doi.org/10.1787/53f23881-en
15. OECD (2023b) PISA 2022 results (II): Learning during — and from — disruption. OECD Publishing. https://doi.org/10.1787/a97db61c-en
16. OECD (2024) *PISA 2022 technical report*. OECD Publishing. https://www.oecd.org/pisa/data/2022database/
17. Pholphirul P, Rukumnuaykit P, Teimtad S (2023) Teacher shortages and educational outcomes in developing countries: Empirical evidence from PISA-Thailand. Cogent Educ 10(2) Article 2243126. https://doi.org/10.1080/2331186X.2023.2243126
18. Presidential Decree of the President of the Republic of Uzbekistan No. UP-79 (2023), May 26 *On measures to further develop the system of teacher certification and improve the qualifications of teaching staff*. https://lex.uz
19. Presidential Decree of the President of the Republic of Uzbekistan No. UP-21 (2026), February 16 *On the Uzbekistan–2030 Strategy*. https://lex.uz
20. Raudenbush SW, Bryk AS (2002) Hierarchical linear models: Applications and data analysis methods, 2nd edn. Sage
21. Rivkin SG, Hanushek EA, Kain JF (2005) Teachers, schools, and academic achievement. Econometrica 73(2):417–458. https://doi.org/10.1111/j.1468-0262.2005.00584.x
22. Rubin DB (1987) Multiple imputation for nonresponse in surveys. Wiley. https://doi.org/10.1002/9780470316696
23. Sanfo JBM (2024) Teaching quality and student achievement inequalities in low-and middle-income countries: A hierarchical linear model analysis. Stud Educational Evaluation 83:101419
24. Sun J, Du Y (2021) Teacher training and the urban-rural education gap in China. Asia Pac J Educ 41(3):463–477. https://doi.org/10.1080/02188791.2020.1790312
25. Woessmann L (2003) Schooling resources, educational institutions and student performance: The international evidence. Oxf Bull Econ Stat 65(2):117–170. https://doi.org/10.1111/1468-0084.00045
26. World Bank (2023) Beyond the learning drop: Why countries in Eastern Europe and Central Asia should act now to avoid a teacher crisis. World Bank Educ Blog. https://blogs.worldbank.org/education/beyond-learning-drop-why-countries-eastern-europe-and-central-asia-should-act-now-avoid