STRENGTHENING THE NATIONAL ASSESSMENT SYSTEM FOR MANAGING EDUCATION QUALITY: INTERNATIONAL PRACTICE AND POLICY PRIORITIES FOR UZBEKISTAN

Dilshoda A. Norboyeva. Republican educational center, Tashkent, Uzbekistan

Abstract. Reliable, well-managed national assessment systems are a precondition for evidence-based education policy, yet many transition economies still lack the institutional capacity to design, administer and use them effectively. This article examines the national system for assessing general secondary education quality in the Republic of Uzbekistan against a theoretical model of national assessment and against the practice of three countries with established systems — Australia (NAPLAN), the United States (NAEP) and Chile (SIMCE) — and against five other members of the Commonwealth of Independent States (CIS). Using comparative and statistical-economic analysis of policy documents, international assessment frameworks and World Bank Human Capital Index (HCI) data, the study finds that Uzbekistan's assessment infrastructure remains fragmented: no single body holds end-to-end responsibility for the assessment cycle, standardized instruments meeting international reliability and validity requirements are largely absent, and the country trained no domestic testing (psychometric) specialists before 2020. These gaps are reflected in outcomes: in its first participation in PIRLS (2021), PISA (2022) and TIMSS (2023), Uzbekistan scored well below international benchmarks, underperforming even the level implied by its 2020 Human Capital Index of 0.61. The article proposes a dedicated National Assessment Center with defined departments for item development, test administration, ICT, data analysis, publishing and psychometric quality assurance, and sets out priority actions for building specialist capacity, regulatory infrastructure and international cooperation. The findings are relevant to education policymakers across Central Asia and to other reforming national assessment systems.

Keywords: national assessment system; education quality management; PISA; TIMSS; PIRLS; Human Capital Index; education policy; Uzbekistan

TA’LIM SIFATINI BOSHQARISH BO‘YICHA MILLIY BAHOLASH TIZIMINI TAKOMILLASHTIRISH: O‘ZBEKISTON UCHUN XALQARO TAJRIBA VA SIYOSAT USTUVORLIKLARI

Annotatsiya. Ishonchli va samarali boshqariladigan milliy baholash tizimlari dalillarga asoslangan ta’lim siyosatini yuritishning zaruriy shartidir; biroq, o‘tish davridagi ko‘plab iqtisodiyotlarda bunday tizimlarni ishlab chiqish, joriy etish va samarali qo‘llash uchun institutsional salohiyat hali ham yetarli emas. Ushbu

maqolada O‘zbekiston Respublikasida umumiy o‘rta ta’lim sifatini baholashning milliy tizimi nazariy model hamda tizimlari yaxshi yo‘lga qo‘yilgan uch mamlakat — Avstraliya (NAPLAN), AQSh (NAEP) va Chili (SIMCE) — shuningdek, Mustaqil Davlatlar Hamdo‘stligi (MDH)ga a’zo yana besh davlat tajribasi asosida tahlil qilinadi. Siyosiy hujjatlar, xalqaro baholash mezonlari va Jahon bankining Inson kapitali indeksi (HCI) ma’lumotlari asosida o‘tkazilgan qiyosiy hamda statistik-iqtisodiy tahlil shuni ko‘rsatadiki, O‘zbekistonda baholash infratuzilmasi hali ham parchalangan holatda: baholash sikli uchun to‘liq mas’uliyatni zimmasiga oladigan yagona organ mavjud emas, xalqaro ishonchlilik va validlik talablariga javob beradigan standartlashtirilgan vositalar deyarli yo‘q va mamlakatda 2020-yilgacha mahalliy testologiya (psixometriya) bo‘yicha mutaxassislar tayyorlanmagan. Ushbu kamchiliklar natijalarda o‘z aksini topdi: O‘zbekiston PIRLS (2021), PISA (2022) va TIMSS (2023) dasturlarida ilk bor ishtirok etganida xalqaro ko‘rsatkichlardan ancha past natijalarni qayd etdi va hatto 2020-yildagi 0,61 ga teng Inson kapitali indeksi darajasidan ham pastroq ko‘rsatkich ko‘rsatdi. Maqolada test topshiriqlarini ishlab chiqish, test o‘tkazish, AKT, ma’lumotlar tahlili, nashriyot ishlari va psixometrik sifatni ta’minlash bo‘limlariga ega maxsus Milliy baholash markazini tashkil etish taklif qilinadi hamda mutaxassislar salohiyatini oshirish, me’yoriy-huquqiy infratuzilmani yaratish va xalqaro hamkorlikni yo‘lga qo‘yish bo‘yicha ustuvor chora-tadbirlar belgilab beriladi. Tadqiqot natijalari Markaziy Osiyodagi ta’lim siyosati ishlab chiquvchilari va milliy baholash tizimlarini isloh qilayotgan boshqa davlatlar uchun muhim ahamiyatga ega.

Kalit so‘zlar: milliy baholash tizimi; ta’lim sifatini boshqarish; PISA; TIMSS; PIRLS; Inson kapitali indeksi; ta’lim siyosati; O‘zbekiston

УКРЕПЛЕНИЕ НАЦИОНАЛЬНОЙ СИСТЕМЫ ОЦЕНКИ ДЛЯ УПРАВЛЕНИЯ КАЧЕСТВОМ ОБРАЗОВАНИЯ: МЕЖДУНАРОДНЫЙ ОПЫТ И ПРИОРИТЕТЫ ПОЛИТИКИ ДЛЯ УЗБЕКИСТАНА

Аннотация. Надежные и эффективно управляемые национальные системы оценки являются необходимым условием для формирования образовательной политики, основанной на фактических данных; однако многим странам с переходной экономикой все еще не хватает институционального потенциала для их разработки, администрирования и эффективного использования. В данной статье анализируется национальная система оценки качества общего среднего образования Республики Узбекистан в сопоставлении с теоретической моделью национальной системы оценки, опытом трех стран с устоявшимися системами (Австралия — NAPLAN, США — NAEP, Чили — SIMCE) и практикой пяти других стран СНГ. На основе сравнительного и статистико-экономического анализа программных документов, международных рамок оценки и данных Индекса человеческого капитала (HCI) Всемирного банка установлено, что инфраструктура оценки в Узбекистане остается фрагментарной: отсутствует единый орган, несущий комплексную ответственность за весь цикл оценки; практически нет стандартизированных инструментов, отвечающих международным требованиям надежности и

валидности; а до 2020 года в стране не велась подготовка собственных специалистов в области тестирования (психометрики). Эти пробелы сказываются на результатах: при первом участии в исследованиях PIRLS (2021 г.), PISA (2022 г.) и TIMSS (2023 г.) показатели Узбекистана оказались значительно ниже международных ориентиров и не соответствовали уровню, ожидаемому исходя из значения Индекса человеческого капитала (0,61) за 2020 год. В статье предлагается создать специализированный Национальный центр оценки с подразделениями, отвечающими за разработку заданий, проведение тестирования, ИКТ, анализ данных, издательскую деятельность и обеспечение психометрического качества, а также определяются приоритетные меры по формированию кадрового потенциала, развитию нормативно-правовой базы и расширению международного сотрудничества. Выводы исследования представляют интерес для лиц, определяющих образовательную политику в странах Центральной Азии, а также для специалистов, занимающихся реформированием национальных систем оценки.

Ключевые слова: национальная система оценки; управление качеством образования; PISA; TIMSS; PIRLS; Индекс человеческого капитала; образовательная политика; Узбекистан

Introduction. Since Becker (1964) and Schultz (1971) formalized human capital theory, education has been treated in economic analysis not merely as a social service but as an investment good whose returns — to individuals, employers and national economies — depend on the quality, not only the quantity, of schooling received. In knowledge-based economies, where labor productivity increasingly rests on cognitive and adaptive skill rather than physical or natural capital, the credibility of a country's human capital claims rests on its ability to measure what students actually learn, and to feed that measurement back into policy and classroom practice. National assessment systems — the institutional machinery for designing, administering, analyzing and disseminating standardized measures of student learning — are the primary tool through which education ministries generate this evidence.

Global engagement with standardized learning assessment has expanded steadily since the mid-1990s. The Trends in International Mathematics and Science Study (TIMSS) grew from 45 participating education systems in its first cycle to 64 by 2019; the Progress in International Reading Literacy Study (PIRLS) grew from 35 systems in 2001 to 60 by 2021; and the Programme for International Student Assessment (PISA) grew from 43 countries in 2000 to 81 countries and economies in 2022. This expansion reflects a broader shift in education governance: policymakers increasingly rely on comparative, standardized evidence — rather than administrative counts of enrolment or attainment — to judge whether reforms are working.

The Republic of Uzbekistan has, since 2017, embedded this logic in a sequence of strategic documents: the Law “On Education” (2020), the Concept for the Development of the Public Education System through 2030, the New Uzbekistan Development Strategy for 2022–2026, and the National Programme for the Development of School Education for 2022–2026. Each commits, in general terms, to bringing the assessment of teaching and learning quality “into line with international

standards.” In practice, however, the institutional architecture for delivering on this commitment has been unstable. The State Inspectorate for Education Quality Control — the body formally responsible for monitoring compliance with state education standards and for coordinating Uzbekistan's participation in international assessment programmes — was abolished on 1 January 2023 under Presidential Decree No. PF-269, with its functions redistributed among ministries and agencies rather than consolidated in a successor body. This leaves an open institutional question at precisely the moment when Uzbekistan is, for the first time, generating internationally comparable evidence on its own learning outcomes.

This article asks two connected questions. First, what does international practice — both in countries with long-established national assessment systems and in Uzbekistan's immediate CIS neighborhood — suggest about the institutional design choices (sampling approach, stakes, reporting practice, staffing) that make a national assessment system function as a genuine management tool rather than a compliance exercise? Second, against that comparative benchmark, what does Uzbekistan's own recent experience — including its first results in PIRLS, PISA and TIMSS — reveal about the priority reforms needed to close the gap between strategic ambition and institutional capacity? The remainder of the article is organized as follows: Section 2 describes the sources and methods used; Section 3 presents the comparative and empirical results; Section 4 discusses their implications and proposes an institutional model for a national assessment center; Section 5 concludes.

Materials and Methods. The study is a comparative policy analysis. Three groups of sources were used. The first comprises Uzbekistan's legal and strategic documents on education quality (laws, presidential decrees and Cabinet of Ministers resolutions issued between 2017 and 2023) together with the corresponding legal and institutional documents of five other CIS states — Armenia, Belarus, Kazakhstan, Kyrgyzstan and Tajikistan — selected because they share Uzbekistan's Soviet-era administrative legacy and post-1991 reform trajectory, which makes their institutional choices a relevant comparator. The second comprises the technical and framework documentation of three national assessment systems selected for their contrasting institutional designs — the National Assessment Program – Literacy and Numeracy (NAPLAN) in Australia, the National Assessment of Educational Progress (NAEP) in the United States, and the Sistema de Medición de la Calidad de la Educación (SIMCE) in Chile — together with the frameworks of the major international comparative programmes (PISA, TIMSS, PIRLS, TALIS). The third comprises quantitative outcome data: the World Bank's 2020 Human Capital Index update, Uzbekistan's own 2019 TIMSS-aligned national learning assessment (conducted to support its first Human Capital Index estimate), and the official results of Uzbekistan's first participation in PIRLS (2021), PISA (2022) and TIMSS (2023), published by the OECD and the International Association for the Evaluation of Educational Achievement (IEA) in 2023–2024.

Methods of grouping, comparison, induction and deduction, and statistical-economic analysis were applied to identify structural similarities and differences across countries and to relate institutional design features to reported outcomes. Because the comparative design relies on secondary policy documents and published

assessment results rather than original data collection, the analysis is best read as a structured synthesis intended to inform policy design rather than a causal evaluation of any single reform.

Results. 3.1 The national assessment system as a management tool: a five-component model

A national assessment can be defined as an exercise designed to describe the achievement of a defined population of students in a given subject area, or areas, at a particular age or grade, using standardized instruments and, typically, questionnaires that collect contextual information (Greaney & Kellaghan, 2008). What distinguishes a national assessment from classroom testing or a school-leaving examination is its purpose: it is built to generate system-level feedback — to ministries, regional authorities and school networks — about whether learning standards are being met, rather than to certify or select individual students.

The literature converges on a five-stage operational cycle that any national assessment system must manage: (i) design, in which the population, grades, subjects and policy questions to be answered are specified; (ii) administration, covering instrument development, sampling or census logistics, and field operations; (iii) data analysis, in which raw responses are scaled, checked and linked to the policy questions defined at the design stage; (iv) reporting, in which results are packaged for distinct audiences; and (v) dissemination and use, through which findings are actually taken up in policy, resource allocation or classroom practice. Weakness at any single stage undermines the value of the whole cycle: an assessment that is well designed and well administered but poorly disseminated produces data with no policy return, while one that is well disseminated but poorly designed produces conclusions that cannot be trusted.

Because reporting needs differ sharply across audiences, mature systems typically produce a tiered set of information products rather than a single report: individual results and practical guidance for parents; item-level diagnostic analysis for teachers; concise summaries with clear implications for school and district administrators; press releases and plain-language briefings for the media; and full technical reports, microdata and online query tools for researchers. Cost planning follows the same logic. In the United States' NAEP programme, for example, data collection typically accounts for roughly 30 percent of total programme cost, instrument development about 15 percent, sampling about 10 percent, data processing about 10 percent, and manual and guideline production about 5 percent, with the remainder covering analysis, reporting and dissemination — a distribution that illustrates why underfunding any single stage, particularly dissemination, can leave the cost of the entire exercise poorly justified.

3.2 International practice: Australia, the United States and Chile

The three systems examined here illustrate genuinely distinct institutional choices rather than variations on a single template. NAPLAN, administered annually since 2008 to all Australian students in grades 3, 5, 7 and 9, is a census-based assessment: every student in the target grades sits the test, individual results are

returned to families, and school-level results are published on the public "My School" website, making the system simultaneously a diagnostic tool for families and teachers and an accountability instrument for the public. NAEP, by contrast, is explicitly designed as a low-stakes, sample-based monitor of national and state trends: schools and students are randomly sampled, no individual student or school ever receives a NAEP score, and the programme's entire policy value lies in tracking system-level trends over time and across states rather than in evaluating any single school, teacher or student. SIMCE, Chile's mandatory national assessment since 1990, occupies an intermediate position: like NAPLAN it is census-based across its target grades (2, 4, 6, 8, 10 and 11), but like NAEP it does not release individual student scores; instead, school-level results are reported and — unlike in the other two systems — are widely used by the media to construct public school rankings, which gives SIMCE stronger accountability "bite" at the school level than either comparator.

Table 1. Comparison of three national assessment systems

Feature	Australia — NAPLAN	United States — NAEP	Chile — SIMCE
Established	2008	Ongoing federal programme; current design from the 1980s–1990s	1988 (voluntary); mandatory under education law since 1990
Coverage	Census: all students in target grades	Sample: randomly selected schools and students	Census: all students in target grades
Grades assessed	3, 5, 7, 9	4, 8, 12 (subjects vary by cycle)	2, 4, 6, 8, 10, 11 (subject and grade combinations vary)
Core subjects	Reading, writing, language conventions, numeracy	Mathematics, reading, science, writing, civics, U.S. history, and others on a rotating cycle	Mathematics, Spanish, history/social studies, natural sciences; English, ICT and physical education added later
Individual	Yes — returned to	No — individual	No — individual

Feature	Australia — NAPLAN	United States — NAEP	Chile — SIMCE
student results	families	scores are never produced	scores are not released
School-level results	Yes — published (“My School” website)	No — no school-level reports	Yes — published and used for media rankings
Primary policy use	Growth tracking and public accountability	National/state trend monitoring; low-stakes by design	Policy evidence base, school accountability, targeted pedagogical support
Governing body	Australian Curriculum, Assessment and Reporting Authority (ACARA)	National Center for Education Statistics, U.S. Department of Education	Agency for Education Quality / Ministry of Education

Source: compiled by the author from national assessment programme documentation.

Despite these differences in stakes and disclosure, the three systems share three design features that recur across mature national assessment systems generally. First, all three supplement test scores with contextual questionnaires — covering student demographics, language background, parental education and employment, and school characteristics — precisely because raw scores are uninterpretable without information on the conditions under which they were produced. Australia's NAPLAN uses an Index of Community Socio-Educational Advantage to cluster schools with statistically similar student intakes before comparing results, a technique that allows “like-for-like” comparison rather than penalizing schools for factors outside their control. Second, all three treat the underlying tests as comparatively low-stakes for individual teachers: even SIMCE, which has the most public accountability exposure of the three, does not attach formal sanctions to low results, relying instead on public disclosure and targeted financial support to underperforming schools. Third, all three have made significant investments in computer-based test delivery — NAPLAN and NAEP both transitioned core subjects to tablet- or computer-based administration between 2017 and 2019 — reflecting a broader international shift that has direct implications for any country building a system from a later starting point.

3.3 Institutional capacity across the CIS: where does Uzbekistan stand?

A comparison of six CIS states — Armenia, Belarus, Kazakhstan, Kyrgyzstan, Tajikistan and Uzbekistan — shows considerable variation in how far national assessment functions have been consolidated into dedicated institutions.

Table 2. Institutional framework for national education-quality assessment in selected CIS countries

Country	Lead ministry	Specialized assessment body	Notes
Armenia	Ministry of Education and Science	Government-affiliated Assessment and Testing Center	The only country in the sample whose legal framework explicitly names the assessment system's purpose — to supply reliable information to learners, parents, teachers, administrators, employers and civil society
Belarus	Ministry of Education	Republican Institute for Knowledge Control	Systematic in-service training pipeline for assessment specialists; supported by a World Bank Institutional Development Fund grant and a 2016 modernization project
Kazakhstan	Ministry of Education and Science	National Testing Center (domestic) and National Center for Education	Dedicated Master's programme in educational measurement running

Country	Lead ministry	Specialized assessment body	Notes
		Statistics and Assessment (international)	since 2014 at the state pedagogical university for women
Kyrgyzstan	Ministry of Education and Science	National Testing Center; Center for Education Assessment and Teaching Methods (independent, U.S.-supported)	Assessment functions split between a state body and a non-governmental organization
Tajikistan	State Control Service in Education	National Testing Center under the Presidency	Regular in-service training for assessment staff; specialists also trained through World Bank-supported international programmes
Uzbekistan	Ministry of Preschool and School Education	No single successor body; the former State Inspectorate for Education Quality Control was abolished 1 January 2023 (Decree PF-269)	No domestic training in testing/psychometrics existed before 2020; a Master's specialization opened at Tashkent State Pedagogical University in the 2020–2021 academic year

Source: compiled by the author from national legal and institutional documents of the countries listed.

The comparison points to a capacity gap rather than a design gap: Uzbekistan's strategic documents describe broadly the same goals as its neighbors, but lack an

equivalent institutional “home” for delivering them. As of the most recent assessment, approximately 500 monitoring staff were distributed across the country's 14 regions, supplemented by roughly 10 specialists at the national education center and 4 at the ministry — yet none had received systematic training in testing theory, psychometrics or data analysis. A two-round training delivered by a single World Bank-financed international expert, while valuable, cannot substitute for the sustained, credentialed training pipelines that Belarus and Kazakhstan have built over the past decade. The absence of a standardized, validated national item bank compounds the problem: the various assessment exercises historically run in Uzbekistan's school system — term and final examinations, entrance tests for specialized schools, subject olympiads — have not been benchmarked against the reliability and validity requirements that international bodies apply, which is one reason none of them has been recognized internationally as equivalent to a national learning assessment.

3.4 Uzbekistan's engagement with international assessments: preparation and outcomes

Between 2020 and 2022, in preparation for its first participation in PISA, PIRLS, TIMSS and TALIS, Uzbekistan's education authorities — working with the OECD, the IEA, the World Bank's READ programme and bilateral experts — undertook a substantial preparatory campaign. This included televised lessons incorporating PISA-style tasks for grades 7–9 (over 3,900 videos with 210 participating teachers), the translation and distribution of methodological guides on PISA-style creative-thinking assessment (500 copies to 348 core schools), teacher training covering roughly 4,500 teachers in mathematical, scientific and reading literacy across three subject strands, a series of international webinars (each drawing 700 to over 1,000 participants) featuring international experts, the creation of the STesting digital platform hosting more than 1,000 international-programme-aligned test items, and the Uzbek-language publication of Andreas Schleicher's “World Class: How to Build a 21st-Century School System” (Schleicher, 2018) in 5,000 copies.
The resulting first-time outcomes, published between 2023 and 2024, are summarized in Table 3.

Table 3. Uzbekistan's results in its first participation in major international student assessments

Programme (cycle)	Domain / grade	Uzbekistan score	Comparator	Rank among participants
PIRLS 2021	Grade 4, reading literacy	437	PIRLS centerpoint: 500	49th of 57

Programme (cycle)	Domain / grade	Uzbekistan score	Comparator	Rank among participants
PISA 2022	Age 15, mathematics	364	OECD average: 472	Bottom range of 81
PISA 2022	Age 15, reading	336	OECD average: 476	Bottom range of 81
PISA 2022	Age 15, science	355	OECD average: 485	Bottom range of 81
TIMSS 2023	Grade 4, science	412	TIMSS centerpoint: 500	54th of 58
TIMSS 2023	Grade 8, mathematics	421	TIMSS centerpoint: 500	32nd of 44
TIMSS 2023	Grade 8, science	396	TIMSS centerpoint: 500	39th of 44

Source: OECD (2023b) for PISA 2022; IEA TIMSS & PIRLS International Study Center for PIRLS 2021 and TIMSS 2023 official country results.

Two features of these results merit particular attention. First, in every programme Uzbekistan's first-time participation placed it in the bottom quartile to bottom decile of participating systems, a pattern consistent with the well-documented “first-participation effect” observed in other education systems entering these programmes for the first time, but also consistent with the structural capacity gaps described in Section 3.3. Second, and more specific to Uzbekistan, the actual TIMSS 2023 results can be compared with the harmonized proxy test score of 474 that Uzbekistan generated in 2019 — through a national assessment of 3,922 grade 5 students in 150 schools, deliberately aligned to the TIMSS 2015 framework — for the specific purpose of estimating its first Human Capital Index. Even allowing for differences in grade level and assessment design between the 2019 proxy and the actual 2023 TIMSS administration, Uzbekistan's subsequently measured performance did not exceed what the earlier proxy implied, and on several indicators fell short of it. This suggests that the 2019 harmonized assessment, however useful as a planning

benchmark, did not overstate the scale of the learning gap subsequently confirmed through full international participation.

3.5 Human capital returns to schooling: the 2020 Human Capital Index

The World Bank's Human Capital Index (HCI) measures the human capital a child born today can expect to attain by age 18, given the prevailing risks to health and education in the country where they are born. It combines indicators of child survival, expected years of schooling, a harmonized test score reflecting the quality of that schooling, and adult survival and stunting rates as proxies for health. Uzbekistan joined the HCI project in 2018 and, in November 2019, conducted the national learning assessment described above specifically to generate an internationally comparable school-quality indicator for its first HCI estimate.

Table 4. Human Capital Index 2020: Uzbekistan compared with regional and income-group averages

Component	Uzbekistan	Europe & Central Asia (average)	Lower-middle-income countries (average)
Probability of survival to age 5	0.978	0.993	0.961
Expected years of school	12.1	13.0	10.4
Harmonized test score (national, TIMSS-aligned)	474	495	391
Learning-adjusted years of school	9.20	10.4	6.6
Adult survival rate (age 15–60)	0.861	0.900	0.807
Human Capital Index (HCI)	0.61	0.70	0.48

Source: World Bank (2020) Human Capital Index database; Uzbekistan country data as reported in Norboeva (2023).

The resulting profile is a nuanced one: Uzbekistan's HCI of 0.61 indicates that a child born in the country today can expect to reach only 61 percent of the productivity they could achieve with complete education and full health, trailing the Europe and Central Asia regional average of 0.70. At the same time, Uzbekistan's index exceeds the average of 0.48 recorded for lower-middle-income countries as a group — the income category in which Uzbekistan was classified at the time — and its overall profile was found to be closely comparable to that of Georgia, Albania and Moldova, countries at similar income levels and in the same broad geographic region. In short, Uzbekistan underperforms its regional peers but outperforms its income peers, a finding that argues against both complacency (relative to the region, the gap is real and the country ranked at the bottom of the field in its first full assessments) and undue pessimism (relative to comparably resourced economies, the starting point is not unusually weak).

Discussion. Read together, Sections 3.3 and 3.4 point to a consistent interpretation: the weak results Uzbekistan recorded in its first participation in PIRLS, PISA and TIMSS are not simply a reflection of classroom-level teaching quality, but the visible output of a national assessment infrastructure that, at the time of testing, lacked a single body with end-to-end responsibility for the assessment cycle described in Section 3.1, lacked a validated domestic item bank benchmarked to international reliability and validity standards, and lacked a trained professional workforce of testologists, psychometricians and data analysts of the kind that Belarus, Kazakhstan and Tajikistan have built through sustained, credentialed training programmes. None of the three benchmark systems examined in Section 3.2 achieved credibility, whether framed as a diagnostic tool (NAPLAN), a low-stakes national monitor (NAEP), or a public accountability mechanism (SIMCE), without first securing exactly this kind of institutional foundation.

Building on the five-component model in Section 3.1 and the comparative evidence in Sections 3.2 and 3.3, a dedicated National Assessment Center — with a clear legal mandate covering the full assessment cycle, rather than functions distributed across multiple ministries and agencies as at present — is proposed as the institutional anchor for reform. A center of this kind would need a professional staff sufficient to cover each stage of the cycle; a working estimate, based on comparable center structures, is approximately 76 core staff organized into the functional departments summarized in Table 5, in addition to technical and support personnel.

Table 5. Proposed functional structure of a National Assessment Center for Uzbekistan

Department / division	Core functions
Test and item development	Develops standardized test and measurement materials aligned with curriculum standards and international frameworks

Department / division	Core functions
Test administration	Plans, organizes and conducts testing; prepares administrator and observer manuals and field procedures
Information and communication technology (ICT)	Develops and maintains the automated (computer- or tablet-based) assessment system
Publishing	Produces manuals, test booklets, answer sheets, questionnaires and public information materials
Organizational-analytical	Builds and maintains the assessment database; manages sampling and response coding
Scientific-methodological expertise and psychometric analysis	Evaluates the reliability and validity of assessment materials; conducts item and scale analysis
Assessment technology	Develops coding guidelines, organizes scoring, and summarizes and synthesizes results for reporting

Source: adapted by the author from the organizational model proposed in Norboeva (2023).

Beyond the organizational chart itself, four priorities follow directly from the comparative evidence. First, capacity-building for assessment specialists needs to move from occasional seminars — such as the two week-long sessions delivered by a single international expert in 2020 — to a sustained, credentialed pipeline of the kind Kazakhstan has run since 2014 and Belarus has embedded in its in-service training system; expanding the Master's-level testology programme launched at Tashkent State Pedagogical University in 2020–2021, including through joint programmes with international partners, is a logical next step. Second, cooperation already initiated with Russia's Center for International Cooperation in Education Development (CICED) on the Student Achievement Monitoring (SAM) instrument, together with continued engagement with the OECD, IEA and World Bank, should be used specifically to accelerate the development of a validated domestic item bank, building on the existing STesting platform rather than starting from scratch. Third, reporting practice should follow the tiered, audience-specific model described in Section 3.1

and observed across all three benchmark countries in Section 3.2, rather than a single undifferentiated report; parents, teachers, school administrators, policymakers and the media each need results presented in a form suited to the decisions they are actually able to make. Fourth, the balance of stakes attached to national assessment results needs to be set deliberately: the evidence from Section 3.2 suggests that a system weighted too heavily toward high-stakes, individually punitive use of results risks distorting teaching toward the test, whereas a system with essentially no public accountability risks the same neglect of dissemination and use that undermines assessment value everywhere it occurs.

The analysis has clear limits. It draws on secondary policy documents and published assessment results rather than original classroom-level or household data, and it cannot establish causal attribution between any single institutional gap and any specific score. Uzbekistan's regional, urban–rural and socio-economic disparities in learning outcomes — dimensions that NAPLAN's socio-educational advantage index and SIMCE's vulnerability index are explicitly designed to capture — could not be examined here and merit dedicated analysis once successive cycles of Uzbekistan's own international assessment data become available.

Conclusion. Uzbekistan's education strategy documents since 2017 have consistently committed the country to aligning its assessment of learning quality with international standards, and its 2020–2023 preparatory campaign for PISA, PIRLS and TIMSS mobilized real resources — teacher training, digital platforms, international partnerships — toward that goal. Yet the comparative analysis presented here shows that strategic commitment has outpaced institutional capacity: no single body currently holds end-to-end responsibility for the national assessment cycle, a validated national item bank benchmarked to international standards does not yet exist, and the professional workforce of testologists and psychometricians remains close to the level it was before 2020. Uzbekistan's results in its first participation in PIRLS (2021), PISA (2022) and TIMSS (2023) — consistently in the bottom quartile to bottom decile of participating systems — are best read not as a verdict on Uzbekistan's classrooms alone, but as the first reliable, internationally comparable evidence of a gap that domestic indicators had not previously been able to show with the same precision.

The comparative evidence from Australia, the United States, Chile and five CIS states converges on a common lesson: a national assessment system functions as a genuine management tool — rather than a periodic compliance exercise — only when it is anchored in a single institution responsible for the full design-to-dissemination cycle, staffed by trained specialists, and equipped with instruments that meet recognized reliability and validity standards. Establishing a National Assessment Center along the lines proposed in Section 4, paired with a sustained specialist training pipeline and continued international cooperation, offers Uzbekistan a concrete path from its current, still-developing HCI position — above its income peers but below its region — toward the human capital outcomes its own development strategy already targets.

Foydalanilgan adabiyotlar:

1. Becker, G. S. (1964). Human capital: A theoretical and empirical analysis, with special reference to education. Columbia University Press.
2. Bolotov, V. A., & Valdman, I. A. (2012). Conditions for the effective use of the results of assessing students' educational achievements. Pedagogika, (6).
3. Greaney, V., & Kellaghan, T. (2008). Assessing national achievement levels in education. World Bank.
4. Greaney, V., Kellaghan, T., & Murray, T. S. (2011). Using the results of a national assessment of educational achievement (National Assessments of Educational Achievement, Vol. 5). World Bank.
5. International Association for the Evaluation of Educational Achievement [IEA]. (2022). PIRLS 2021 international results in reading. TIMSS & PIRLS International Study Center, Boston College.
6. International Association for the Evaluation of Educational Achievement [IEA]. (2024). TIMSS 2023 international results in mathematics and science. TIMSS & PIRLS International Study Center, Boston College.
7. Mullis, I. V. S., & Martin, M. O. (Eds.). (2019). PIRLS 2021 assessment frameworks. TIMSS & PIRLS International Study Center, Boston College.
8. Norboeva, D. A. (2023). Improving the national assessment system in education quality management [Master's dissertation, Higher School of Business and Entrepreneurship under the Cabinet of Ministers of the Republic of Uzbekistan].
9. Organisation for Economic Co-operation and Development [OECD]. (2018). Education at a glance 2018: OECD indicators. OECD Publishing. https://doi.org/10.1787/eag-2018-en
10. Organisation for Economic Co-operation and Development [OECD]. (2023a). PISA 2022 results (Volume I): The state of learning and equity in education. OECD Publishing.
11. Organisation for Economic Co-operation and Development [OECD]. (2023b). PISA 2022 results (Volume I and II) – Country notes: Uzbekistan. OECD Publishing. https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbcc5-en/uzbekistan_2bb94bf1-en.html
12. President of the Republic of Uzbekistan. (2022a). On the development strategy of New Uzbekistan for 2022–2026 (Decree No. PF-60, 28 January 2022).
13. President of the Republic of Uzbekistan. (2022b). On approval of the national programme for the development of school education for 2022–2026 (Decree No. PF-134, 11 May 2022).
14. President of the Republic of Uzbekistan. (2022c). On measures to implement administrative reforms of New Uzbekistan (Decree No. PF-269, 21 December 2022).
15. Republic of Uzbekistan, Cabinet of Ministers. (2018). On measures to organize international research in the field of education quality assessment in the public education system (Resolution No. 997, 8 December 2018).
16. Republic of Uzbekistan. (2020). Law of the Republic of Uzbekistan "On Education."

17. Schleicher, A. (2018). World class: How to build a 21st-century school system. OECD Publishing.
18. Schultz, T. W. (1971). Investment in human capital. Free Press.
19. World Bank. (2020). The Human Capital Index 2020 update: Human capital in the time of COVID-19. World Bank.