

Developing a national assessment that supports Kazakhstan's education goals

Introduction

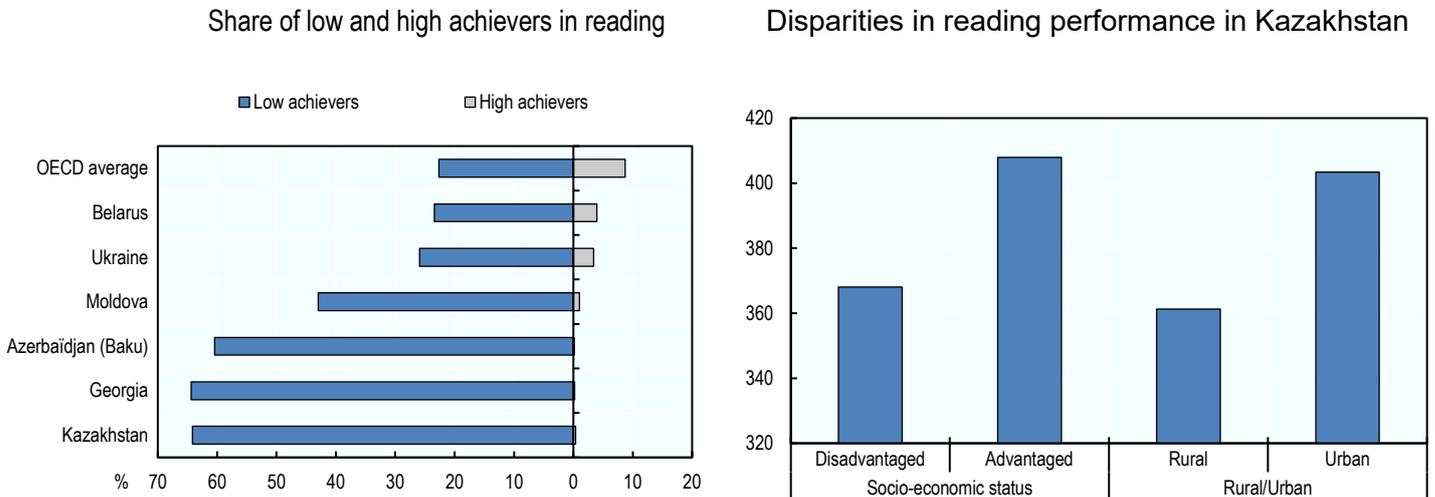
Since achieving independence in 1991, Kazakhstan has undergone rapid development and emerged as a regional economic leader. Recent ambitions, expressed notably in the long-term strategy *Kazakhstan 2050*, aim to strengthen and diversify the economy in order to position the country as a global leader (Republic of Kazakhstan, 2016^[1]). A key priority in this strategy is to develop the knowledge and professional skills of Kazakhstan's population, 28% of whom were below the age of 15 in 2018 (compared to an OECD average of 18%) (World Bank, 2018^[2]). Kazakhstan has already made tremendous progress in providing access to all levels of schooling. Today, enrolment in primary and lower-secondary education is nearly universal. Moreover, almost all graduates from lower-secondary school continue to either general upper-secondary school or vocational studies and roughly half of Kazakhstanis between the ages of 25 and 34 now hold a tertiary degree, which is greater than the OECD average of 41% (see Annex A).

Having achieved high levels of educational access, Kazakhstan is now turning its attention towards improving educational quality. To understand progress in this area, Kazakhstan benchmarks its educational performance against those of leading economies through international surveys, such as the OECD's Programme for International Student Assessment (PISA). Results from PISA 2018 reveal that the average Kazakhstani student scored around 100 points below the OECD average in reading and around 64% of Kazakhstani students were unable to achieve a baseline level of reading proficiency needed to participate fully in society (OECD, 2019^[3]). This share of low performers is much higher than the OECD average (23%) and one of the highest among PISA participating countries in the OECD Eurasia Competitiveness Programme (Figure 1).

Results from PISA 2018 also show large degrees of inequity in Kazakhstan. Factors such as socio-economic background and, in particular, school location can influence students' performance (Figure 1). Whether the schools of Kazakhstani students are in rural or urban communities explains a greater share of student variance in reading performance (6.7%) than across OECD countries (4.5%). For instance, students in Nur-Sultan city, the national capitol, scored 428 on average, compared to 344 for students from Atyrau, a comparatively more rural region (Figure 2). These findings can be partially explained by a national focus on developing a cadre of very high-achieving students combined with a lack of adequate attention to improving education provision in marginalised areas. In 2008, the government established the Nazarbayev Intellectual Schools (NIS), a network of 20 high-performing schools to which entrance is highly selective and competitive. While students from these schools achieve impressive outcomes, the pedagogical initiatives they have incubated are difficult to scale and not always well adapted to schooling environments in all parts of the country. Meanwhile, students in areas such as Atyrau struggle to achieve

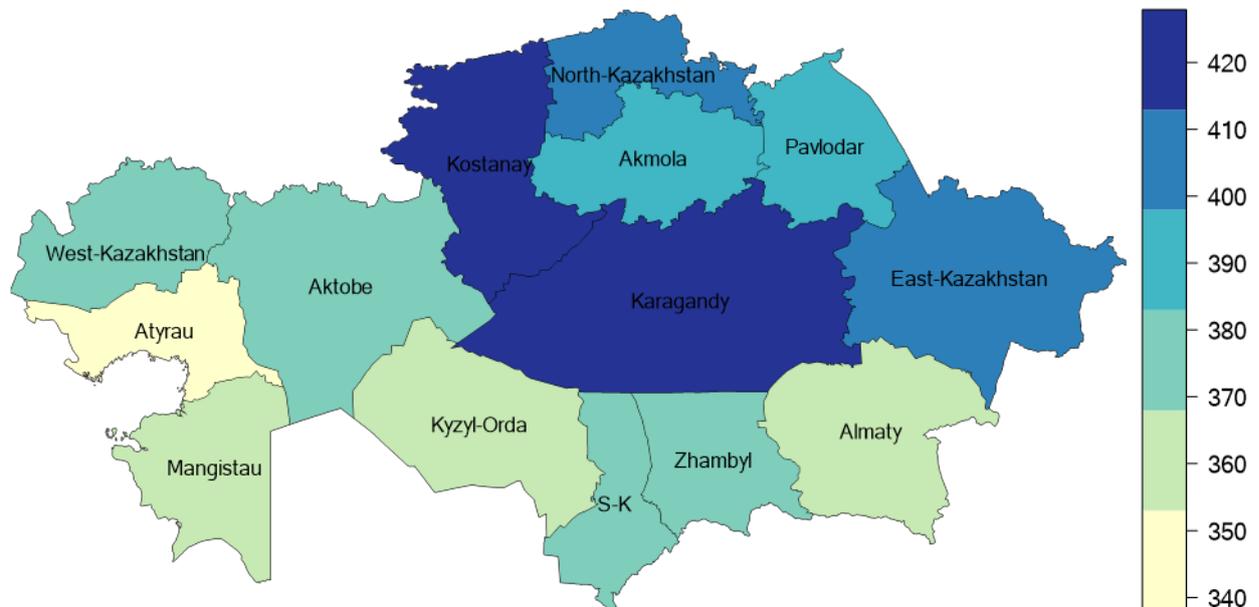
basic minimum standards, influencing their chances of attending tertiary education and finding good employment.

Figure 1. Reading performance in Kazakhstan in PISA 2018



Note: The 13 countries included in the OECD Eurasia Competitiveness Programme are Afghanistan; Armenia; Azerbaijan; Belarus; Georgia; Kazakhstan; Kyrgyzstan; Mongolia; Republic of Moldova; Tajikistan; Turkmenistan; Ukraine and Uzbekistan. Only countries with PISA data are included in the figure.
 Source: PISA 2018 Database.

Figure 2. Regional differences in reading performance



Note: Not depicted are the cities of Nur-Sultan and Almaty, which have special administrative status. Nur-Sultan scored 428, while Almaty scored 424.

The challenge of widening inequalities in Kazakhstan is compounded by demographic trends that are straining the system's capacity to provide a quality education for all students. Rapid urbanisation has created overcrowded schools in cities across the country. As of 2018, over 6% of students attended schools that operated in triple shifts (IAC, 2019^[4]). Meanwhile, achieving universal access to education in a large country with many remote communities has created an extended network of small rural schools that face challenges related to poor infrastructure and staff shortages (IAC, 2019^[4]; OECD/The World Bank, 2015^[5]). Particularly representative of these circumstances are “ungraded schools”, which do not have enough students to form full classes of separate grades. As of 2018, around 41% of public schools were ungraded schools, though they only enrolled 6% of the student population (IAC, 2019^[4]).

To develop the sustainable and knowledge-based economy that Kazakhstan envisions, the government needs to create systems and instruments that help it understand how all students are performing and how they can be supported in their learning. This OECD country review examines four educational policy areas (see Box 1) that Kazakhstan can focus on in order to improve the outcomes of all students.

Box 1. The OECD's review of education evaluation and assessment policies in Kazakhstan

This policy perspective is one in a series of four that draw on an OECD knowledge-base created through reviews of evaluation and assessment policies in over 25 education systems. To complete this review, the Ministry of Education and Science of Kazakhstan (hereafter, the ministry) and the OECD review team chose a specific policy issue within four broad areas of evaluation and assessment (student assessment, teacher appraisal, school evaluation and system evaluation). The selected issues are:

- Strengthening national examinations in Kazakhstan to achieve national goals
- Raising the quality of initial teacher education and support for early career teachers in Kazakhstan
- Developing a school evaluation framework to drive school improvement
- Developing a national assessment that supports Kazakhstan's education goals

The review of these policy issues was based on national information that Kazakhstan provided to the OECD, background research and a visit to different parts of the country in November 2019. During the visit, a team of OECD staff met with key actors across the education system to discuss the policy issues. This evidence formed the basis of the policy perspectives, each of which provides actionable recommendations based on insights from international practices to help Kazakhstan strengthen student learning while making learning outcomes more equitable.

The importance of national assessments

National assessments are standardised tests that represent one of the most important tools education systems have to monitor student learning and track progress towards national education goals. Unlike examinations, national assessments do not have consequences on students' progression or certifications. Instead, their primary purpose is usually to provide reliable data on student learning outcomes for system monitoring (OECD, 2013^[6]). National assessments can also serve other purposes, such as providing information to schools and teachers to help enhance student learning and/or supporting school accountability or evaluation frameworks. Moreover, when accompanied by background questionnaires, national assessments offer valuable insights into the factors influencing learning across a country and specific groups of students.

Key features of the national assessment system in Kazakhstan

Kazakhstan has a well-established national assessment system, the External Assessment of Academic Achievement (EAAA), which measures student learning at the end of each curriculum cycle. However, there are growing concerns that the assessment's current design and processes for disseminating and using results do not effectively support system evaluation nor help the country to achieve its national education goals. As a result, the ministry is considering a reform proposal to eliminate the EAAA and replace it with a new national measure of student learning that will possibly be called Monitoring of Students Educational Achievements (MSEA).

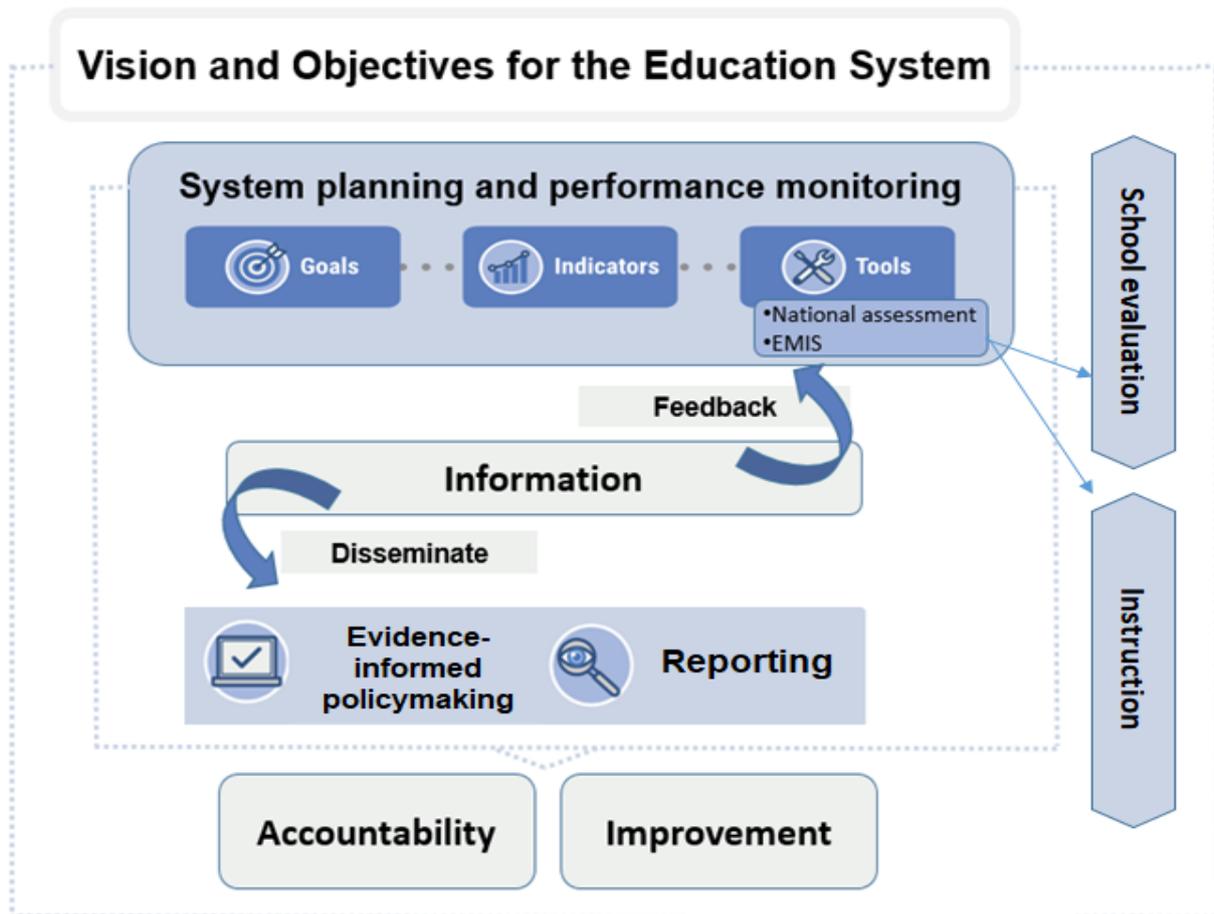
This policy perspective provides timely recommendations for harnessing the political appetite to improve Kazakhstan's national assessment system and develop it into a tool that not only supports system monitoring but also helps to evaluate and improve the education sector. The following policy perspective first focuses on the role that the EAAA plays in evaluating Kazakhstan's education system overall. It then discusses other potential purposes for the assessment, namely how it can more effectively support policymaking, school evaluation and teaching and learning. Using a single assessment to serve different purposes has important implications for its design and carries risks that need to be fully understood and managed. Therefore, this policy perspective also addresses the EAAA's key design features, offering suggestions on how to adapt the instrument so that it can better support Kazakhstan's education goals.

Providing information for system evaluation

System evaluation refers to the processes that countries use to monitor and evaluate the performance of their education system as a whole (OECD, 2013^[6]). The two main functions of system evaluation are to hold the education system, and actors within it, accountable for achieving their stated objectives; and, to help improve educational processes. System evaluation has gained increasing importance in recent decades across the public sector, in part because of growing pressure on governments to demonstrate the results of public investment and improve efficiency and effectiveness (Schick, 2003^[7]). This is especially true in the education sector, which by ensuring that all students master foundational literacy and numeracy skills, has a central role in preparing children for success in a knowledge-based society and increasing economic competitiveness.

Countries use a range of information sources to evaluate the education system and track progress towards national goals. One vital source of information comes from national assessments (see Figure 3), the designs of which vary considerably across countries (OECD, 2015^[8]). However, there is a consensus that having regular, reliable national data on student learning is essential for both system accountability and improvement. Today, the vast majority of OECD countries (30), and a growing number of non-Member countries, now have regular national assessments of student achievement (OECD, 2015^[8]). To understand the EAAA's role in supporting system evaluation in Kazakhstan, it is important to recognise some of the key features of system evaluation in general and how the EAAA relates to those features.

Figure 3. System evaluation



National vision

System evaluation needs to be steered by a national vision and/or goals, which provide reference points against which to evaluate performance. In many countries, these goals are set out in an education strategy that spans several years. Kazakhstan has a long-term strategic vision to become one of the 30 most developed countries in the world by 2050 and there is a regular cycle of medium-term plans to help achieve this goal. Most recently, the ministry adopted Kazakhstan's 2020-25 State Program for the Development of Education and Science (hereafter, State Programme for Education) in December 2019.

While Kazakhstan's previous State Programmes for Education focused mainly on ensuring access, the current strategy focuses more on equipping all students with the competencies needed to develop a competitive economy and cohesive society. Notably, the 2020-25 State Programme recognises the importance of improving the quality of external student assessments, to provide objective and equitable information about student learning (Republic of Kazakhstan, 2019^[9]). This shift towards concentrating on outcomes is positive. It also highlights the timeliness of this policy perspective since the EAAA is an important tool for monitoring the outcomes-based strategy and collecting data that will help advance the country's broader goals to improve educational equity and quality.

Indicators

Indicators are the quantitative or qualitative variables that help to monitor progress towards national goals, while targets clearly express a desired level of performance (World Bank, 2004^[10]). Countries use different

types of variables in their indicator frameworks to measure progress, including *inputs* like government spending, *outputs* like teacher recruitment, and *outcomes* like student learning. While outcomes are notoriously difficult to measure, they are a feature of frameworks in most OECD countries because they help measure the results that a system is trying to achieve (OECD, 2009^[11]). The European Union, for example, set a benchmark of having no more than 15% of 15-year-old students score below Level 2 across PISA mathematics, reading and science tests. This goal has influenced many national targets, with some countries aiming to go even further, such as Ireland, which set a target of having no more than 10% of low-achieving students across PISA tests (Department of Education and Skills, 2016^[12]).

Kazakhstan's State Programme for Education has a detailed list of indicators to help monitor and evaluate the achievement of national goals. For each indicator, the ministry has identified a source of information, baseline data on the current condition of the indicator and targets to benchmark the desired outcomes midway and at the end of the programme's duration. The 2020-25 State Programme for Education uses data from international assessments and the EAAA as key indicators. For example, there is a target using PISA and TIMSS data to measure the goal of reducing gaps in educational quality between rural and urban regions. There are also targets of mean scores on EAAA tests to measure the goal of implementing an updated quality assessment system. Reporting these types of commitments support system accountability and transparency; however, the State Programme for Education does not disaggregate EAAA data beyond mean scores to set targets intended to reduce achievement disparities.

Tools

Governments can draw on various tools to collect data for monitoring and evaluation. Most countries have established at least two main types of data collection in the education sector. The first is administrative information about students, teachers and schools, such as enrolment, human resources and school facilities. Countries often store this data centrally in their Education Management Information Systems (EMIS). The second type of data collection focuses on learning outcomes, which are available through tools such as national and international assessments, as well as national examinations. Kazakhstan has several well-developed tools that collect information about the country's education system.

National databases

The National Education Database (NED) is a sophisticated tool that electronically collects and stores Kazakhstan's administrative data for all levels of education, from early childhood to higher education. Other databases hold different types of information, such as student grades and results on national assessments. At present, Kazakhstan is working to integrate the NED with these other education databases using common student identification numbers. Such links will enhance the analytical functions of national education data. The Information Analytic Centre (IAC) also recently developed an online data portal, called the System for Education Data Analysis (SEDA), which allows users to visualise and explore some of the publically available data in the NED. Currently, the SEDA platform does not contain EAAA results.

National assessment

The EAAA is Kazakhstan's main national tool for measuring student learning at the system level. To understand the current design of this instrument, it is important to acknowledge the historical context of Kazakhstan's national assessment system. Prior to the EAAA, Kazakhstan's Quality Control Committee introduced the Interim State Control (ISC) assessment in 2005 to help evaluate and improve school quality (Bridges, 2014^[13]). Schools that scored poorly on this census-based assessment risked closure, which contributed to cheating and test manipulation, undermining the assessment's value as a monitoring tool. The accountability burden and growing mistrust of the ISC's results led to its abolishment by presidential order in 2011 (World Bank, 2012^[14]).

Kazakhstan’s experience with the ISC had important implications for the purpose, design and use of the EAAA, which replaced the ISC in 2012. For example, the EAAA is intentionally sample-based to reduce the risk of its association with direct stakes for schools, teachers or students. Table 1 provides key information about the EAAA today and the remainder of this section discusses selected features in greater detail. Despite Kazakhstan’s efforts to address some of the shortcomings and challenges of its previous national assessment (ISC), there are increasing concerns that the EAAA is no longer fit for purpose. In particular, the assessment does not align with the country’s new competence-based curriculum and the results it produces are not widely trusted. As a result, the EAAA is not used to inform policy making nor help improve teaching and learning practices.

The ministry is currently planning reforms to improve or replace Kazakhstan’s national assessment system. In the upcoming months, policy makers will take important decisions regarding the EAAA’s core purposes, design features and the dissemination and use of results. This ongoing national assessment reform coincides with a unique opportunity for Kazakhstan to develop a more modern assessment system through funding and technical support provided by the World Bank Education Modernisation Project. Specifically, the World Bank project includes an activity to improve Kazakhstan’s external student assessment system in an effort to raise learning outcomes, reduce disparities among rural and disadvantaged schools, and support system monitoring and evaluation more generally (World Bank, 2017_[15]).

Table 1. Key information about Kazakhstan’s national assessment, EAAA

Topic	Summary		
Stated purposes	Monitoring students’ academic achievement Assessing the effectiveness of teaching and learning Conducting a comparative analysis of the quality of educational services provided by schools		
Grade and frequency	Annual assessments in Grades 4, 9 and 11		
Target population	Sample-based, representative of 20% of schools in each region (<i>oblast</i>), schools included on a rotating basis		
Testing mode	Tests taken on computer or paper, depending on available school infrastructure		
Subjects	<i>Grade 4</i> (est. 2016)	<i>Grade 9</i> (est. 2012)	<i>Grade 11</i> (est. 2017)
Time	2 subjects selected by Quality Control Committee 70 minutes	1 mandatory subject (Kazakh language): number and type of other subject domains vary, determined annually by Quality Control Committee 130 minutes	3 subjects selected by Quality Control Committee 140 minutes
Score	Max score: 30 points	Max score: 75 points	Max score: 100 points
Item type	Close-ended: multiple choice (single answer or multiple answer)		
Variables collected	<ul style="list-style-type: none"> No student background survey for Grade 4; student background survey has 18 questions for Grades 9 and 11; collects information on gender (boys/girls), language of instruction and shift of school; there are proxies for student socio-economic background 14 questions in school administrator background survey; collects information on school socio-economic characteristics, teacher qualifications, shortages of subject teachers, textbooks used and material resources 		
Marking	Regional testing centre offices mark tests electronically and submit results to the National Testing Centre’s central office		
Results	The National Testing Centre reports results aggregated at rayon, oblast and national levels; individual school-level results are also available on the National Testing Centre’s (NTC) public website; some analysis of results are disseminated in reports. Participating schools receive school-level results within three days		

Source: Data provided from the Ministry of Education and Science

- Grades and frequency

Kazakhstan currently administers the EAAA at the end of the year to students in Grades 4, 9 and 11, which correspond to the end of the primary, lower secondary (known in Kazakhstan as basic secondary), and

upper-secondary education cycles (resulting in a 4+5+2 curriculum model). Within the context of the ongoing national assessment reform, the ministry is considering dropping the assessment in Grade 11 and moving administrations from the end of Grades 4 and 9 to the end of Grades 5 and 10. The latter change would only take place if adding Grade 12 to the school structure, which is currently being discussed by the ministry, results in extending curriculum cycles for primary and basic secondary education (i.e. moving to a 5+5+2 curriculum model). Together, these changes aim to reduce overlap between the national assessment and exams (for Grade 11) and continue providing a measure of student learning at the end of the primary and lower-secondary curriculum cycles.

- Target population

At present, the EAAA covers no more than 20% of schools that are representative at the regional (oblast) level. Schools are included in the sample on a rotating basis so that each school will participate in the EAAA at least once every five years. This design provides an important means to compare student achievement in different parts of the country and helps reduce the risk of associating the assessment with stakes. However, the arrangement simultaneously prevents the generation of timely and reliable information about individual students and schools, which can be useful for identifying areas of low-performance and planning interventions to inform policy and support pedagogy.

- Testing mode

Kazakhstan's National Testing Centre (hereafter, the NTC) and the World Bank, under the Education Modernisation Project, are helping to improve the school facilities and technical infrastructure for assessment. These changes will allow more Kazakhstani schools, especially those in rural areas, to administer standardised tests electronically. Currently, this option is only available for schools with sufficient computer and Internet access. In terms of the EAAA, schools can choose to administer the assessment via paper or computer. Regardless of the delivery mode, all EAAA tests (even paper versions, which are scanned) are marked electronically. This approach allows participating schools to receive their results within three days after testing, helping to raise public trust in the marking process. However, offering the EAAA in both computer and paper formats could lead to mode effects, meaning there is potential for students to receive different results because they took different versions of the same test. The presence of mode effects would jeopardise the EAAA's value as a system-monitoring tool.

- Tested subjects

The Quality Control Committee annually determines what subjects are included in the EAAA (with the exception of Kazakh language, which is mandatory on the Grade 9 EAAA). The Grade 4 EAAA includes two subjects and the Grade 11 EAAA includes three subjects. The Committee also determines how many and what non-mandatory subject domains are included in the Grade 9 EAAA. Importantly, the full sample of Grade 9 students do not take all of the non-mandatory subjects each year, rather tests covering different subject areas are evenly distributed to students participating in the sample. This approach means that in a given year, some Grade 9 students might be tested in Algebra and Physics while others are tested in the History of Kazakhstan, Biology and Informatics. While this approach aims to cover all areas of the curriculum over time, it reduces the sample size for each subject area in the Grade 9 EAAA and undermines the reliability of national trend data to help monitoring progress towards education goals.

- Item types

The EAAA has low construct validity, meaning the tests do not adequately measure Kazakhstan's new competence-based curriculum and learning standards. For example, the literacy test asks students to select epic genres of Russian literature from a given list of different writing forms, rather than testing their reading comprehension. This type of question reflects students' ability to memorise and recall knowledge rather than apply critical thinking skills to answer a question. The ministry is currently working with the World Bank to help improve the quality of content and items on the EAAA and UNT by training test-developers to introduce new and open-ended questions (see policy perspective on examinations).

This approach aims to complement exclusively multiple-choice items and measure a wider range of skills, including important higher-order thinking skills that more closely reflect the new curriculum. Introducing such item types implies major changes to the way the EAAA is marked and in the Kazakhstani context, using human markers risks fuelling concerns about assessment reliability.

- Variables collected

The EAAA currently has background questionnaires that students and school administrators complete when participating in the assessment sample. The information collected from the student questionnaire (only for Grades 9 and 11) allows results to be disaggregated by gender, language of instruction and geographic area (urban/rural), among other things. There are also proxies for student socio-economic background, such as parental level of education and resource availability in the home (e.g. a home library, Internet access, etc.). The EAAA school administrator questionnaire provides valuable information about teachers' qualifications, subject areas facing teacher shortages and the quality of textbooks and resources available in the school. More contextual information will soon become available when the EAAA results database is linked with official administrative data from the NED. Together, the EAAA background questionnaires collect robust information about factors associated with student performance in Kazakhstan. However, some questions risk becoming redundant when the EAAA and NED databases are integrated. Moreover, Kazakhstan does not fully exploit data from either survey to develop pedagogical interventions and inform education policy.

International assessments

Kazakhstan began participating in large-scale international assessments of student achievement in 2007, starting with the Trends in International Mathematics and Science Study (TIMSS) and then the Programme for International Student Assessment (PISA) in 2009 (see Table 2 for a full list). Today, Kazakhstan is one of the few members of the Commonwealth of Independent States (of the former Soviet Union) that participates broadly and consistently in international education assessments and surveys, reflecting an increasing openness to scrutiny and comparison. International assessments like TIMSS and PISA support system evaluation by allowing Kazakhstan to benchmark teaching and learning practices and outcomes with other countries. Since there have been many changes to Kazakhstan's national assessment system in recent years, international assessments also provide an important source of trend data about the education system. However, these tools cannot help measure the extent to which students are achieving the national curriculum and learning standards.

National examinations

Kazakhstan has two national examinations that provide data with the potential to support system monitoring. The two exams are the UNT, which serves as an entrance examination for entry into university and the Final Attestation, which serves as a school-leaving exam in Grade 11. The policy perspective on national examinations covers these exams in greater depth but the instruments can also serve as system monitoring tools, albeit with some limitations. In particular, data from the Final Attestation is not fully standardised and only students in general upper-secondary schools take the exam, excluding nearly 40% of students at this level who attend vocational schools. On the other hand, the UNT exam changes content from year to year, meaning that this test does not allow for reliable comparisons over time. The UNT also has limited use in assessing student achievement in relation to national standards at the system level because it does not align to the curriculum nor does it include a representative sample of students. Kazakhstan, however, plans to introduce new items types into the UNT in 2021.

Disseminating information collected by system evaluation tools

Once information has been generated by system evaluation tools, countries must decide how to disseminate that information (OECD, 2013^[6]). Reports are the most common dissemination mechanism and can cover a range of topics such as results from national assessments, annual statistics, or specific policies. Many countries produce a specific report on national assessment results, in addition to feeding this information into broader system-wide reports. The main national assessment report serves as a basis for communicating findings to the minister and senior education officials, parliament and the media, which in turn can support policy discussions and evidence-informed decision-making. If national assessment data is to be useful to teachers and schools, it is also important to analyse results from an instructional perspective (e.g. by identifying common student errors). To ensure that assessment data meets the variety of needs and interests of diverse audiences, countries often generate different types of reports for different actors, including students, teachers, schools, and policy makers (Kellaghan, Greaney and Murray, 2009^[16]). Other methods of disseminating information include making the raw data available to researchers and developing interactive portals that allow users to manipulate data easily.

In addition to the format in which information is disseminated, countries must decide what information to make available publically or only to a limited audience. To this end, governments may widely disseminate reports or choose send them privately to select actors. Similarly, data can be shared online for everyone to access or only be available by request. Many OECD countries have committed to strengthening the transparency of their education sector, which is essential for well-functioning evaluation and assessment systems and often results in greater ownership and acceptance of reforms among stakeholders (OECD, 2013^[6]). However, the extent to which assessment data is made public and the rate at which transparency increases depends on country context and how information is likely to be used and interpreted. For example, systems with a history of high-stakes testing that are trying to lessen the pressure around tests might consider withholding school-level assessment results from the public since the extra transparency could generate negative pressure that might lead to test manipulation.

Kazakhstan's annual *Report on the State and Development of the Education System* draws on EAAA data, using it as part of an index on the effectiveness of sub-national education systems (IAC, 2019^[4]). The NTC also publishes EAAA data on its website in the form of PDF and Excel files that include the aggregate scores and names of individual schools participating in the sample. Notably, there is no report that specifically analyses EAAA data. The implications of how - and to what extent - Kazakhstan disseminates and uses EAAA data publically are discussed below in detail.

Evidence-informed policy making

To justify the effort and expenditure involved in implementing a national assessment, the information it provides should be useful to policy makers, curriculum developers, textbook writers, teacher trainers, and the public (Kellaghan, Greaney and Murray, 2009^[16]). Leveraging the potential of a national assessment to support reform and drive system improvement requires having the capacity to analyse the data and connect findings in relevant and meaningful ways to the work of different actors. For example, ministries should understand assessment data and be able to use it to inform planning decisions.

In Kazakhstan, the education ministry and its technical agencies sometimes evaluate specific areas of interest to inform policy decisions. For example, the IAC conducted a study in 2016 on the organisation of homework that led to a ministerial order calling for a reduction in the amount of time students spend doing schoolwork at home (MoES, 2019^[17]). Despite this promising example, Kazakhstan - like most countries - does not have a systematic approach to using evaluation information to inform policy making, especially when it comes to using national assessment data. This is partly because the EAAA’s limited coverage rate and the lack of comprehensive analysis of its background questionnaires are currently hindering the assessment’s ability to provide policy makers with timely and relevant insights.

Supporting teaching and learning

National assessments are unlikely to improve teaching and learning unless the findings are used to improve school and classroom practices (Kellaghan, Greaney and Murray, 2009^[16]). For example, insights from national assessments can feed into school improvement processes, but also initial teacher education and professional development programmes. When assessments are census-based, they can help track and support student learning by providing an external reference for teachers to moderate and benchmark their classroom grading, which can help strengthen the quality and equity of teacher’s professional judgements about student achievement. To this end, many countries develop specific materials (e.g. item analysis, guidelines and trainings) to support teachers and schools in interpreting national assessment results and using them to enhance student learning.

Kazakhstani schools that participate in the EAAA sample receive their results as a list of students’ aggregate raw scores by subject and grade level. There are very limited benchmarks for comparing schools and there is no analysis of EAAA data from a pedagogical perspective, such as materials identifying which parts of the curriculum students tend to struggle with and might need more attention in the classroom. While regional Departments of Education can conduct their own analysis of EAAA results for pedagogical purposes, there are no examples or guidelines on how to do this. As a result, the current approach to disseminating EAAA data leaves schools and teachers without robust contextualised information, analysis and support to devise strategies aimed at improving learning outcomes.

Informing school evaluation

Some countries use national assessments to help evaluate school quality. Research suggests that using a single indicator, such as a school result on a national assessment, is not an accurate indication of a school or teacher(s) effectiveness since it does not consider contextual factors outside of the school’s control (OECD, 2013^[6]). The practice of publishing school-level results or using them to issue rewards or sanctions can lead to the perception that assessments have stakes. This perception, in turn, can have unintended and negative consequences, such as test manipulation or influence on student enrolment distributions (e.g. parents may choose to send their children to schools with high results, which could lead to oversubscription in some schools). However, when countries use assessment results in a balanced way and integrate results data with other school-level information, they can help benchmark school performance more equitably and meaningfully. This practice can help governments better identify and understand some of the factors behind low performing schools (or groups of students). Shanghai (China),

for example, has developed a “green indicator framework,” which includes assessment results as one of ten school performance indicators, alongside others such as student health, socio-economic background, workload and the pedagogical practices used in the school (among others).

Kazakhstan has a history of using national assessment results punitively, as part of the former school control system that was widely perceived as unfair and subject to interference. This approach led to concerns about the assessment's integrity and the reliability of results. Today, the EAAA is sample-based and therefore cannot be used inappropriately as a single indicator of school quality. However, the ongoing national assessment reform is considering plans to once again provide school-level results of student learning outcomes. This information can be a valuable tool to help Kazakhstani education officials identify and address areas of low performance, especially considering the challenges of having such an extensive school network. However, changing the EAAA's design to produce school-level results will require consideration for the country's previous experience with full-cohort assessments. In particular, it will be important to ensure that results are used as part of a robust set of indicators for making comprehensive and contextualised comparisons of school quality.

Agencies with responsibilities for developing and/or using assessment

In many countries, standardised assessments are managed by a specialised agency with some degree of independence. This autonomy *vis á vis* education authorities is important to help ensure that technical judgements are not influenced solely by political opinions. While education ministries can certainly lead decisions about the national assessment, they need input from a range of actors to shape the design and direction of the assessment system effectively. As a result, countries often establish clear governance arrangements to oversee important decisions related to their national assessment.

National Testing Centre develops and implements the national assessment instrument

In addition to its central office, the NTC operates out of 154 regional offices located throughout the country, which together are responsible for developing and managing all of Kazakhstan's national assessments and exams, in particular the Unified National Test (UNT) and the EAAA. In the past, however, the Quality Control Committee (see below) managed the national assessment. The decision to have a technical body manage the EAAA reflects Kazakhstan's efforts to build capacity and expertise in implementing large-scale assessments of student learning. The present allocation of responsibilities also aligns with how many OECD countries manage their assessment systems.

The NTC has staff with a range of skill sets and experiences, including statisticians and psychometricians. As a result, this agency is well placed to design and implement national tests and analyse the data they produce. However, the NTC faces challenges in terms of recruiting software developers and IT specialists, which could present a problem as Kazakhstan works to modernise its national assessment system. For example, a shortage of IT specialists may hinder the NTC's ability to benefit from opportunities created by computer-based test delivery. The World Bank Education Modernisation Project plans to train IT professionals to help address this challenge.

Despite the NTC's technical expertise and responsibility for developing and implementing national tests, the agency sees its role as executing decisions taken by the ministry and the Quality Control Committee. In particular, the NTC lacks an independent remit to engage systematically in reform discussions about how to improve the national assessment instrument. This prevents the EAAA from developing in a way that is technically feasible while trying to meet the evolving needs of the education community.

Quality Control Committee is leading the national assessment reform

The Quality Control Committee (hereafter, the Committee) is a subordinate unit within the ministry that sets and carries out quality control policies in education and science. While the Committee no longer has

responsibilities for designing and managing Kazakhstan’s national assessment system, it still controls the integrity of the EAAA’s administration. In 2018, this meant sending Committee staff to 1406 different schools to serve as test monitors for each participating school in the EAAA sample. This practice presents a capacity concern since the Committee only has around 357 staff members within its central and regional offices located across the country. Controlling the EAAA’s administration so tightly also risks reinforcing the perception that the assessment carries stakes.

Importantly, the Committee is one of the primary users of EAAA data, which it relies on as one of the core indicators for external school audits (known as preventative control). As a result, the Committee is leading reform efforts to expand the EAAA’s coverage. While it is important for this body to be involved in steering the assessment, broader representation is needed to ensure the EAAA’s design supports Kazakhstan’s national education goals, not only the Committee’s goal of having a standardised external measure of learning outcomes for individual schools. Moreover, the Committee has a longstanding reputation as a regulating and control body, rather than an educational one. If the Committee is seen to be the main actor driving the national assessment reform, this could negatively shape public perceptions of the EAAA.

Information Analytic Centre (IAC) uses assessment data for national reporting

One of Kazakhstan’s most prominent education agencies involved in system evaluation is the Information Analytic Centre (IAC). With around 72 staff members, many of whom have skills in quantitative and qualitative analysis and the language proficiencies needed to work in Kazakhstan’s multilingual environment (Kazakh/Russian/English), the IAC has a high level of technical capacity to carry out system evaluation. This agency also manages Kazakhstan’s participation in large-scale international assessments related to education and the NED. As a technical agency, the ministry often commissions the IAC to conduct independent studies and evaluations. For example, the IAC produces Kazakhstan’s annual Report on the State and Development of the Education System, which includes analysis of EAAA results. In this way, the IAC uses EAAA data to support transparency and provide a source of information to help inform planning and policy making decisions.

Sub-national authorities are important users of assessment data

Kazakhstan is a very large country with an extensive school network. As a result, territorial authorities take many decisions about how to support schools under their responsibility. This context generates a need for data at the sub-national level. Currently, the EAAA covers a school sample that is representative at the regional (oblast) level but does not permit analysis of districts (rayons). The latter therefore depend on (unreliable) Final Attestation data and (only partially representative) UNT data to determine the extent to which students are achieving the national learning standards in individual schools. This situation offers limited information for local authorities to inform their planning and policymaking decisions.

Since the capacity of sub-national authorities to use evaluation results may vary significantly, many countries centrally analyse data to support regions in understanding and using assessment information (OECD, 2013^[6]). In Kazakhstan, oblasts receive a list of schools in their region that participate in the EAAA sample with raw aggregate scores and no contextual information. Fortunately, the IAC’s State of Education report offers some comparative analysis across regions and large cities with special administrative status. While this report shows trends across regions and over time, the analysis mainly highlights areas with the best and worst performance, failing to consider resource and capacity differences. There is also no accountability framework or expectation for regional and district education departments to report what they are doing to address low performance. This disconnect reduces pressure to improve learning outcomes and makes it difficult for the central ministry to implement reforms, such as the new curriculum.

Review of the context

It is important for Kazakhstan to continue investing in a variety of tools and activities to support system evaluation and improvement of the education sector. This policy perspective focuses on one such tool, the national assessment, which plays a key role in generating reliable data on learning outcomes across different student populations, school contexts and over time. Kazakhstan's current national assessment, the EAAA, helps monitor student achievement at the system level. However, there is substantial evidence that the design of the assessment and the dissemination and use of its results make the EAAA unable to support its other purpose of evaluating (and improving) the effectiveness of teaching and learning and comparing the quality of educational services in schools.

One of the most notable challenges facing Kazakhstan's national assessment system is the lack of alignment with national education and development goals. For example, the 2020-25 State Programme for Education places a strong emphasis on strengthening national capacity for innovation but the EAAA does not ask questions that require students to use creative thinking and higher-order analytical skills. The State Programme for Education also recognises the importance of improving the academic achievement of all students and schools, regardless of their socio-economic status (Republic of Kazakhstan, 2019^[9]). However, there is very limited analysis of the EAAA's background questionnaires to help understand the factors that affect teaching and learning to help improve educational equity.

As part of a broader effort to update country's overall evaluation and assessment system, the Kazakhstani ministry is currently considering several initiatives to reform the purpose, design and use of the EAAA. This reform has potential to expand perceptions of student achievement beyond success in examinations and academic competitions towards a broader vision that more closely aligns with national goals. For this to happen, Kazakhstan will need to increase the EAAA's coverage to produce school-level results.

A full-cohort national assessment would generate information about the extent to which individual students are meeting national learning standards, helping to strengthen the quality and equity of teachers' classroom grading. It would also allow education authorities to more accurately identify and support struggling schools, inform education policy and practices (e.g. amend the curriculum), and provide input to sector-planning decisions, such as how to reorganise the school network and address problems of ungraded schools. While school-level results have a clear value, they also carry risks, especially in contexts like Kazakhstan's where there is a tradition of high-stakes testing. To avoid distorting the focus on helping all students to learn, it will be important to reflect carefully about how to disseminate and use EAAA results.

The following recommendations provide timely suggestions on how to harness the political appetite for improving Kazakhstan's national assessment system and developing it into a tool that not only monitors student achievement but also collects and mobilises evidence to help inform policy-making and support teaching and learning. To this end, the EAAA needs strong and consistent leadership. At present, the Quality Control Committee - which was responsible for Kazakhstan's previous assessment system - is the main driver of the EAAA reforms, despite having limited technical expertise in designing robust national assessments and supporting teaching and learning. Without a more diverse governance structure, the EAAA will continue to lack a strategic direction and struggle to fulfil its core purposes.

Recommendation 1. Strengthen governance of the national assessment to refine and prioritise its purposes

Kazakhstan has nearly 15 years of experience in developing and administering national assessments of student achievement. This tradition has contributed to a general awareness within government and among education stakeholders about the value of having a national assessment system. However, the EAAA does not have a distinct role within Kazakhstan's overall assessment framework as there are eight different tests

that have a primary purpose to support system monitoring. Moreover, the EAAA’s leadership is not inclusive nor is there a clear mechanism for considering technical expertise when making decisions about the assessment instrument. As Kazakhstan plans reforms to its educational assessment system, it has become clear that the EAAA needs a strong and inclusive governance structure if it is to support system-wide improvement and inform education policies and practices. Refining and prioritising the key purposes of the EAAA can help guide subsequent design and dissemination decisions to establish an assessment instrument that is better fit for purpose.

1.1. Establish a steering committee to refine the national assessment’s purpose and provide leadership for its strategic development

Evidence

Education ministries and their reform objectives typically guide national assessment systems, but how they do so can vary. Many countries set up dedicated steering committees to help determine the strategic direction, purpose and subsequent design of a national assessment. While the composition and organisation of these steering committees vary, they often include high-level representatives from the ministry of education, especially policy analysts and curriculum bodies, agency(ies) responsible for the assessment’s implementation, teachers, and teacher educators (Greaney and Kellaghan, 2012^[18]). This composition helps establish the assessment’s credibility in the eyes of key stakeholders and can help align the assessment with curriculum reforms, evaluation processes and other education policy goals.

In Kazakhstan, senior members of the education ministry participate in decision-making processes about the implementation and use of the assessment. However, the Quality Control Committee, which served as custodian of the previous national assessment, plays a leading role in today’s ongoing reform of the EAAA. While the Committee is an important user of assessment data, its mandate as an audit agency and the general perception of its role as legal regulators, present a significant risk that schools and the broader public may associate a new Committee-led EAAA as having high stakes. The Committee also lacks the technical expertise to make decisions about the assessment’s design and is not well placed to develop a test that can inform education policies and practices to support system-wide improvement.

While it is positive the Committee has consulted with the IAC and the Centre for Pedagogical Measurement (which monitors student progress in Nazarbayev Intellectual Schools) about the EAAA reform, these consultations are organised as ad-hoc meetings and neither the ministry’s strategic planning committee nor the Y. Altynsarin National Academy of Education (which is responsible for developing the new competence-based curriculum), seem systematically engaged in reform discussions. Moreover, there is no explicit space for the NTC to provide technical input into the reform decision-making process (see Recommendation 1.2). Without more structured and inclusive governance processes, it will be difficult for the new EAAA to measure outcomes that can help evaluate Kazakhstan’s national education goals and enable the type of accountability desired by the government.

Recommended actions

Establish a steering committee with representative leadership

To strengthen Kazakhstan’s national assessment system, this review recommends establishing a high-level steering committee to lead the ongoing EAAA reform and the assessment’s strategic direction, purpose and subsequent design. Similar to other committees under the remit of the ministry of education, the EAAA steering committee should be led by a senior official (i.e. deputy minister) but should also include representatives from other educational agencies who can systematically give advice on the tests’ measurement properties, use in research and educational value. While actors who have already been

involved in discussing the EAAA reform, such as the Quality Control Committee, should sit on this new steering committee, a wider range of actors should also be included. In particular, the heads of the National Academy of Education and the NTC, should provide their respective expertise on curriculum and test instruments to the decision-making process. By establishing more representative leadership of the EAAA, the steering committee can help promote a better understanding of what the assessment data can (and cannot) reveal and be used for. It can also help defend the assessment's validity when results are released, ensure adequate financial support and co-ordinate the efforts to implement the assessment instrument.

Establish clear working processes for the new steering committee

The steering committee will likely need to have frequent meetings until decisions about the current EAAA reform have been decided and implemented. However, it is important that meetings continue on a regular basis later on - not only according to ad-hoc needs – in order to ensure structured leadership and strategic direction. To this end, the ministry might establish a five-year mandate for the steering committee, which in turn should convene once or twice per year – i.e. one month after the release of results - to take stock of the situation, discuss challenges and make any necessary changes to the instrument. Publishing the list of individuals and agencies who sit on the steering committee and the decisions they take (e.g. online, in a declaration or meeting summary), can help promote greater transparency and build support for the national assessment so that its results are trusted and used effectively.

1.2. Support the National Testing Centre's capacity to develop and manage the EAAA

Evidence

It is crucial that whatever agency is responsible for implementing a country's national assessment has the required technical competence. In many countries, technical assessment agencies often operate with some degree of political and financial independence to ensure the integrity of their evaluations (OECD, 2013^[6]). For example, Mexico's Ministry of Education provided autonomy to the National Institute of Education Evaluation (*Instituto Nacional para la Evaluación de la Educación*)¹ in 2013 to lead assessment and evaluation efforts, including national and international student assessments (INEE, n.d.^[19]).

In Kazakhstan, the NTC is responsible for implementing the national assessment. A subordinate unit of the ministry, the NTC has around 450 staff members with a range of skill sets and experiences, including statisticians and psychometricians. Despite its technical expertise and responsibility for developing and implementing the EAAA (and national exams), the NTC is not well positioned to play a strategic planning role in developing the national assessment. While the NTC makes ad-hoc proposals about the format and structure of national tests, their lack of independence may hinder Kazakhstan's ability to systematically adapt measures of student learning with what is technically feasible to better meet the education community's evolving needs.

Recommended actions

The above Recommendation and Recommendation 4 of the policy perspective on examinations argue that the NTC should more systematically contribute to decisions about national examinations and assessments by sitting on the respective steering committees and establishing a management board with representative leadership to help develop and co-ordinate all of Kazakhstan's national tests (UNT, Final Attestation and EAAA). Specifically, the NTC director should sit on the new EAAA steering committee and actively participate in discussions about the purpose and design of the national assessment.

¹ Mexico's INEE was dismantled by a constitutional reform in 2019

To fulfil its broader mandate, the NTC might also receive some political and financial independence so that it can fully develop its specialised expertise. For example, the NTC should be able to pay for training or hire staff with the technical profiles needed to analyse test results for different audiences and express any concerns it may have about changes requested by other members of the management board. The NTC should also develop systematic ways to draw on the knowledge and experience of other education stakeholders with assessment experience. In particular, IAC staff have received training on coding and data processing, project implementation and data analysis and reporting through managing Kazakhstan’s participation in large-scale international assessments of student learning. Failure to share this knowledge with NTC staff represents a missed opportunity for Kazakhstan’s national assessment system to develop in line with global practices.

1.3. Consider making formative feedback a distinct purpose of the EAAA within the national assessment framework

Evidence

National assessments can serve a variety of purposes. Generally, their primary purpose is to monitor system performance; however, they can also be used (among other things) to support school improvement and inform teaching and learning practices. Since fulfilling different purposes requires different design decisions, it is important that assessment systems explicitly define and prioritise the purposes of each test (Newton, 2007^[20]). Internationally, such alignment is usually achieved through the development of national assessment frameworks (see the policy perspective on examinations), which then act as a reference point to help determine the appropriate design characteristics of an assessment.

Kazakhstan does not have a national assessment framework that sets out the distinct purposes of each standardised assessment across the system. The lack of a coherent framework contributes to important purposes not being fulfilled by any test. For example, the government centrally administers ten large-scale assessments of student learning, eight of which have a primary purpose to support system monitoring and two which certify and select students (see Table 2). Importantly, none of these tests are explicitly designed to help teachers teach and students learn.

As a national assessment, the EAAA is best positioned to serve a formative role in Kazakhstan’s assessment system. However, its three stated purposes – to monitor students’ academic achievement, assess the effectiveness of teaching and learning and compare the educational quality of schools – do not include improving educational practice (MoES, 2019^[17]). Consequently, neither the EAAA’s design nor the materials associated with its results fully exploit the instrument’s formative potential. In particular, only a sample of students take the test, and the data are not used to produce reports or pedagogical materials that explain the results’ implications on teaching and learning practices.

Table 2. Large scale, standardised tests of student learning in Kazakhstan

Schooling level	Grades	Assessment type	Body responsible	Frequency	Population	Primary purposes
Primary education (Grades 1-4)	Grade 4	TIMSS (international assessment)	Information Analytics Centre (IAC)	Four-year cycle	Sample	System monitoring
	Grade 4	EAAA (national assessment)	National Testing Centre	Annual	Sample	System monitoring
	Grade 4	PIRLS (international assessment)	Information Analytics Centre (IAC)	Five-year cycle	Sample	System monitoring
Lower-secondary	Grade 8	TIMSS (international assessment)	Information Analytics Centre (IAC)	Four-year cycle	Sample	System monitoring
	Grade 8	ICILS (international	Information Analytics Centre	Five-year	Sample	System monitoring

education (Grades 5-9)	assessment)	(IAC)	cycle			
Grade 9	EAAA (national assessment)	National Testing Centre	Annual	Sample	System monitoring	
Grade 9/10 (age 15)	PISA (international assessment)	Information Analytics Centre (IAC)	Three-year cycle	Sample	System monitoring	
Upper secondary education (Grades 10-11, soon to include Grade 12)	Grade 11 (soon to be Grade 12)	Final Assentation (exam)*	National Testing Centre**	Annual	Census	Certify completion of upper-secondary education
	Grade 11 (soon to be Grade 12)	Unified National Test (exam)***	National Testing Centre	Annual	Voluntary*	Select students into university
	Grade 11	EAAA (national assessment)	National Testing Centre	Annual	Sample	System monitoring

Notes: *Only students who complete general upper-secondary education take the Final Attestation, which accounts for around 60% of students at this level.

**The NTC develops the test; however, schools are responsible for granting students with the Final Attestation certificate.

***Only students who wish to enter university take the UNT.

Students in Kazakhstan also take an examination at the end of Grade 9 to certify their completion of lower-secondary education. Students then determine if they wish to enter general or vocational upper-secondary schooling. However, this test is not centralised (it is developed and administered locally) and is thus not included in this table nor discussed in this policy perspective.

Source: Data provided by the Ministry of Education and Science.

Recommended actions

This review recommends that the EAAA steering committee consider prioritising formative feedback as the EAAA's distinct purpose (alongside its system monitoring purpose) within Kazakhstan's national assessment framework. The process of refining and prioritising the EAAA's key purposes should be part of Kazakhstan's broader efforts to improve the co-ordination of all national tests (see Recommendation 1 in the policy perspective on examinations). This approach would differentiate the main functions of the EAAA in relation to other large-scale assessments administered in the country, help navigate trade-offs and orient design features towards the most important purposes (Newton, 2007^[20]).

Prioritising system monitoring and formative feedback as the EAAA's core purposes (its other uses will be discussed later) can also help address Kazakhstan's evident need for a timely, external assessment of student learning to help improve teaching and learning practices in hard to reach schools. The following recommendations in this policy perspective address key decisions the steering committee will need to take in order to align the EAAA's design with its newly refined purposes. In particular, if Kazakhstan wants the EAAA to inform education policy and support system-wide improvement, the assessment must generate data that can help strengthen pedagogy, curriculum implementation and quality of schooling. To this end, the Kazakhstan should expand the EAAA's coverage to collect data at the school and student level and ensure that all stakeholders use results appropriately.

Recommendation 2. Establish a formative census-based assessment at the end of primary and lower-secondary education

Once the steering committee has refined and prioritised the key purposes of the national assessment, it will need to take several decisions regarding its design. Aligning the EAAA's design with a more formative purpose has the potential to expand the focus of education stakeholders beyond the success of individual high-performing students towards helping all children to learn and succeed. For this to happen, Kazakhstan will need to expand coverage of the EAAA to produce student-level results that can be used to directly support teachers and schools. School-level findings can also inform education policy and sector-planning

decisions. The following recommendation aims to guide deliberations about the feasibility and desirability of different aspects of the EAAA’s design, with the intention that it support a prominent formative function.

2.1. Establish a formative census-based assessment at the end of primary and lower-secondary education

Evidence

Countries can choose to administer national assessments to some students, or all students. Sample-based assessments are favorable if the primary purpose of the assessment is to provide information for system monitoring and related policy making. However, they cannot be used to understand the performance of all students and schools, information which could be used to improve teaching and learning (Greaney and Kellaghan, 2008^[21]). On the other hand, a census-based approach collects information from all students, but is more time consuming and costly to administer. While both sample and full-cohort assessments can have formative value, countries typically determine the target populations for their national assessment system based on their specific contexts and needs. Many countries use a combination of approaches to support different purposes, alternating between census- and sample-based assessments across different grade levels and subjects.

Over time, a growing number of OECD and partner countries have started to administer full-cohort assessments in the early years of schooling. There are several reasons for this trend, notably a desire to ensure that all students master foundational competencies. Some countries have also adopted census-based national assessments to support school accountability efforts; however, there tends to be much more variation and controversy related to using external assessments for this purpose because it risks detracting from the assessment’s formative value. In Chile, for example, the national assessment system was believed to put too much emphasis on accountability, so a 2016 reform led to a reduction in the number of census tests (Paulo Santiago et al., 2017^[22]). Today, Chile only administers annual full-cohort assessments in Grades 4 and 10² which provide reliable measures of student performance to help allocate resources and inform school evaluations.

In Kazakhstan, the EAAA is currently sample-based, which produces aggregate information for system monitoring but prevents most teachers from using the data to better understand the performance of their students and benchmark classroom grading. As a result, the EAAA is unable to help address issues of grade inflation nor directly support teachers in improving their instructional and assessment practices, such as adopting criterion-based assessment practices, which compares student performance to national standards rather than their classmates (a practice referred to as “norm-referenced assessment”).

The lack of school-level data also limits the EAAA’s ability to inform sector-planning decisions and has implications for its reliability as an indicator for internal and external school evaluation. Kazakhstan’s potential “school review” process, for example, plans to rely on a range of information –including EAAA results- to determine which schools are performing well and which require additional support (see the policy perspective on school evaluation for an in-depth discussion of school review). However, without annual EAAA results for each school, the process will not have a timely and objective measure about student learning for most schools.

² Chile also administers a census-based assessment every two years in Grades 6 and 8, in alternate years. Additionally, there are sample-based assessments in select subjects and grades throughout compulsory education.

Recommended actions

In the short term, this review recommends that the EAAA become an annual census-based assessment with a clear formative purpose, administered at the end of the primary and lower-secondary curriculum cycles. The OECD (2014^[23]) made a similar recommendation in a previous review of secondary education in Kazakhstan. While generating student-level results each year provides only a snapshot of student learning in specific curriculum areas and at a specific time, this information can help improve the quality of classroom assessments by giving teachers direct examples of how their students perform against learning standards, while helping identify students in need of additional support. As educators benefit from the insights of a census-based EAAA, this would in turn help reinforce other national goals of implementing a competence-based curriculum and criterion-based assessment. Finally, having an annual full-cohort assessment at the end of primary and lower-secondary curriculum cycles would position the EAAA to better inform sector-planning, support the potential school review process, and provide insights for conducting internal school evaluation.

In the medium to long term, as student performance improves and Kazakhstan's school evaluation system matures, the EAAA steering committee might re-evaluate the country's need for standardised assessment data of individual students. Until then, rapidly filling the absence of comparable information about learning outcomes at the student and school level is an important step towards improving educational equity and supporting overall system development.

Maintain a sample-based assessment in the Grade 11 EAAA

In regards to the target population for the Grade 11 EAAA, the OECD recommends maintaining this as a sample-based assessment. While full-cohort assessments can measure the knowledge and skills of students as they graduate from secondary school, the costs associated with this design outweigh the potential benefits. Notably, a census assessment at this level would have lower formative value compared to earlier grades because teachers would receive results much too late to inform instruction and address potential learning gaps of students. This approach would also increase the testing burden on students, many of whom will take the United National Test and the Final Attestation exam during the same year, which can provide some school-level information. As a result, Kazakhstan should concentrate on strengthening the school-leaving exam at this level of education (see policy perspective on examinations), rather than expanding coverage of the Grade 11 EAAA.

2.2. Implement the census-based assessment in pencil format and gradually develop the infrastructure for delivery via computer*Evidence*

Computer-based assessment offers many advantages compared to pencil and paper formats. For example, it tends to be cheaper to administer (after initial capital investment), is less prone to human error and integrity breaches, and delivers results more quickly. For these reasons, the use of computers to administer national assessments is becoming more common, particularly in countries that introduced a national assessment recently (OECD, 2013^[6]).

The NTC's regional offices already mark the EAAA electronically, which allows participating schools to receive their results within three days after testing (MoES, 2019^[17]). The Kazakhstani government has also placed a strong emphasis in recent years on equipping schools with computer hardware, software, and Internet connection. Efforts to modernise educational resources align well with *Digital Kazakhstan*, a broader national programme launched in 2017 to improve economic competitiveness and quality of life for citizens through the progressive development of the country's digital ecosystem (Government of Republic of Kazakhstan, 2017^[24]). Building on this momentum, Kazakhstan plans to start shifting the delivery of

some subjects in the Grades 9 and 11 EAAA to a fully computer-based format. A regulation that young students in Kazakhstan should not spend more than 25 minutes with digital devices within one study hour, makes it impossible to currently administer the 70 minute Grade 4 EAAA as a fully computer-based test.

Another important factor driving Kazakhstan’s ambition to scale up computer-based testing is the financial and technical support available through the World Bank’s Education Modernisation Project. At the time of this review, the project’s student assessment focus aims to improve Kazakhstan’s technological infrastructure so that a greater number of schools participating in the EAAA’s sample (for Grades 9 and 11) can administer tests via computer.

Despite the broad commitment and support for transitioning to an exclusively computer-based national assessment, Kazakhstan continues to face several challenges. While schools can already choose to administer the EAAA online in Grades 9 and 11, many still lack updated and working computers. There are also frequent Internet connectivity problems that hinder testing processes. According to national data, most schools (98.5%) have broadband access as of 2018, but this share was only 70% for schools located in rural areas. As a result, the NTC reports that only 24% of schools in 2018 used computers to administer the EAAA. Similar issues were present during Kazakhstan’s participation in PISA 2018, which marked the country’s first time administering PISA entirely by computer. In fact, some schools had to borrow computers from other schools in order for all students included in the sample to take the PISA test.

Recommended actions

It is unlikely that Kazakhstan will be able to deliver a computer-based test to a full-cohort of students in the short term. As a result, this review recommends that Kazakhstan continue making the EAAA available in either pencil and paper or computer-based format, which will ensure that all schools can participate in the EAAA, regardless of their technological infrastructure.

While Kazakhstan’s priority should be to collect information about student learning at the student-level, this approach does not prevent the government from gradually developing the infrastructure for nation-wide computer-based assessment. To this end, delivering the Grade 11 assessment, which this review recommends remain sample-based, exclusively via computer could be a feasible medium-term (i.e., 3-5 year) goal. This strategy would allow Kazakhstan to maintain the political emphasis on modernising the education assessment system without limiting the formative potential of the EAAA to help teachers benchmark the performance of their students and support school evaluation processes. Gradually administering national assessments exclusively via computer would also give the ministry more time to evaluate the system’s readiness for computer-based assessment, address remaining issues and organise a communications campaign to prepare schools, teachers, parents and students for this change.

2.3. Take steps to ensure comparability between paper and digital formats of the national assessment

Evidence

When using two modes of the same assessment, with the intention of comparing results, it is important that the digital mode mimic the paper version to the greatest extent possible. However, reviews of the literature reveal that changes in testing modes can sometimes contribute to changes in performance, meaning there is still potential for students to receive different results because they took the paper or computer version of the same test (Bürger, Kroehne and Goldhammer, 2016^[25]). While it is hard to determine the exact reasons for differences in student performance across test modes (these might include variances in students’ familiarity with computers, for example), understanding the extent to which test format affects student performance is critical if the results will be used to monitor learning over time and compare differences across the country.

In Kazakhstan, there have been no in-depth investigations as to the effect of testing mode on student performance in the EAAA. It is therefore unclear if results between the paper and computer versions of the test can be compared reliably. Since urban schools are more likely to administer the EAAA via computer than schools located in rural areas, the potential consequences of mode effects could be significant when using the data to compare performance across regions and geographic locations.

Recommended actions

Allowing schools to choose between the pencil- or computer-based formats of the EAAA ensures that teachers and students are familiar and comfortable with the testing mode and that all schools are able to participate in the assessment. However, it is important that these two formats yield comparable results. To this end, Kazakhstan should conduct research (the IAC is well placed to do so) about the effects of the test mode on student performance. Such actions are especially important if Kazakhstan is to use the EAAA to track achievement over time and between different population groups. If studies determine that test mode effects are present in Kazakhstan, the country may need to adjust the EAAA's scaling methodology to ensure reliability.

2.4. Adjust quality assurance measures to administer a census-based assessment

Evidence

While national assessments do not carry stakes for students, some stakeholders may perceive these instruments as having consequences, especially when results are used to issue rewards or sanctions for teachers or schools. This perception can lead to problems in test administration, such as teachers helping students answer questions or failure to stick to time limits. To help ensure the quality of test administration, countries can implement a test administration form (completed at the end of the testing session) for proctors to provide a record of the extent to which administrative processes were followed (Greaney and Kellaghan, 2012^[18]). Some countries also organise unannounced visits by quality control personnel to oversee the administration in a sample of participating schools.

Kazakhstan currently sends members of the Quality Control Committee to serve as test monitors for each school participating in the EAAA sample. While stakeholders argue that such measures are necessary to prevent cheating and ensure schools and students take the assessment seriously, controlling the EAAA's administration so tightly also risks reinforcing the perception that it carries stakes. Moreover, this approach represents a significant capacity concern if Kazakhstan hopes to expand the EAAA's coverage to a census (as recommended by this review), since the Committee only has around 357 staff members located across the country.

Recommended actions

Quality assurance measures are important, especially in Kazakhstan where national assessment results have a history of being associated with direct consequences for schools. As a result, it will likely take time before Kazakhstani schools fully understand the refined purpose and value of the EAAA and can reliably administer the assessment with high levels of integrity. In the meantime, this review recommends that Kazakhstan adjust current quality assurance measures since the Quality Control Committee will not have enough staff members or resources to oversee the administration of the EAAA in each school once the assessment becomes census-based. Greaney and Kellaghan (2012^[18]) suggest some measures that Kazakhstan should consider:

- *Test administration form.* For each testing session, the test administrator should complete a form to record the extent to which proper administrative procedures were followed, record the start and end times for each section of the test, and signal any special circumstances or

problems that occurred during the administration (e.g. unclear instructions, lack of time problems with testing materials, etc.).

- *Sample of unannounced quality control visits.* Instead of requiring the Quality Control Committee to monitor all schools participating in the EAAA, the Committee might select a sample of schools to undergo quality control visits to ensure the quality of administration. While all schools should know that a possibility exists that they will be monitored, in practice only around 10-20% of schools would undergo such visits. Monitors should also complete a form to report the conditions of administration during their visits.

2.5. Maintain end of year testing schedule at the end of each curriculum cycle and consider introducing a centrally-developed diagnostic test in earlier grades

Evidence

The grade level and time of year that countries decide to administer standardised tests of student learning reflect their national context and the purposes of a particular assessment. Common considerations include at what points students have completed curriculum cycles, when other important activities or assessments occur, at what stage results are most valuable to help teachers identify and address low performance and when results can inform a student's decision about the future of their education. For example, many OECD countries with upper-secondary examinations tend not to administer national assessments at the same time to avoid duplicating efforts and overburdening students.

At present, Kazakhstan administers the EAAA on an annual basis at the end of key curriculum cycles (i.e., the end of Grades 4, 9 and 11). International assessments, like PISA and TIMSS, also provide information about student learning in Grades 4 and 9, as do the national exams (the UNT and Final Attestation) in Grade 11. However, international assessments are only administered once every three to four years, are sample-based and do not measure the extent to which students are mastering the national curriculum. On the other hand, Kazakhstan's UNT exam is not representative of all students (only those who wish to enter university) nor is the Final Attestation (only includes students in general secondary). Moreover, the marking of the Final Attestation is not standardised. As a result, neither exam is well placed to support system monitoring at the upper-secondary level. Importantly, Kazakhstan also has no comparable data about student learning outcomes prior to Grade 4, when children are around ten years-old.

Within the context of ongoing national assessment reforms and plans to extend compulsory schooling to Grade 12, the Kazakhstani education ministry is considering two proposals related to when it administers the EAAA:

1. Eliminating the Grade 11 assessment, starting in 2021, to reduce the testing burden on students, many of whom also take national exams at the end of upper-secondary school.
2. Moving the lower grade assessments from Grades 4 and 9 to Grades 5 and 10, to maintain an assessment at the end of each curriculum cycle. This change would only take place if adding Grade 12 to the school structure results in a 5+5+2 model that extends curriculum cycles for primary and lower-secondary education to Grades 5 and 10, respectively.

While both proposals present valuable considerations, more could be done to ensure better complementarity with Kazakhstan's overall national assessment framework.

Recommended actions

Because Kazakhstan wishes to continue using the EAAA to measure student learning the end of key curriculum cycles, this review recommends the following arrangement:

Maintain full-cohort assessments at end of each curriculum cycle

This review recommends that Kazakhstan maintain a testing schedule that assesses students at the end of the primary and lower-secondary curriculum cycles. Under the current school structure, the EAAA and international assessments, such as TIMSS (in Grade 4) and PISA (in Grade 9/10), are administered in the same grades every three to four years. While excluding schools that participate in international tests from the EAAA sample helps reduce the testing burden on schools and students, Kazakhstan could do more to further ensure complementarity among large-scale student assessments while the current school structure is still in place. For example, the EAAA should be distinguished within the country's overall national assessment framework (see Recommendation 1) by developing the instrument's pedagogical value. To this end, Kazakhstan should align the EAAA with the new curriculum to ensure that core competences can be measured and share results of individual students and schools in a constructive and informative manner (see Recommendation 3) to make the most of having a full-cohort assessment. These measures should be taken regardless of whether or not the extended school structure results in lengthening curriculum cycles to Grades 5 and 10.

Importantly, Kazakhstan should ensure the full-cohort EAAA does not carry stakes for students. While the risk of EAAA results having consequences on students at the end of primary school (Grades 4/5) is minor, student-level results at the end of lower-secondary school (Grade 9/10), could be more consequential. Students entering upper-secondary education must transition into either general or vocational and technical tracks, which could lead to some educators and families using student EAAA scores to direct low achieving students to vocational tracks. Currently, Kazakhstani students decide which upper-secondary programme to pursue based on several factors, such as their likelihood being able to afford higher education. Another key factor is the Grade 9 Final Attestation, a locally developed examination not covered in these policy perspectives because it is not a large-scale standardised assessment. In the future, Kazakhstan might consider replacing the Grade 9 Final Attestation with the census-based Grade 9 EAAA. As a standardised test, the Grade 9 EAAA (turned exam) would provide the externality needed to establish a more equitable measure to help orient students to different education tracks. However, it would be important that students continue considering a range of factors to make decisions about their educational futures and that general upper-secondary schools are not allowed to set entry criteria based on results.

Establish a unique role for the Grade 11 EAAA within the overall student assessment framework

Implementing several standardised tests of student learning and using the results effectively requires considerable capacity and financial resources. It is therefore positive that the Kazakhstani ministry is reviewing whether or not the Grade 11 EAAA is necessary, especially since the country already has information about learning outcomes in Grade 10 from the sample-based PISA assessment, and in Grade 11 (soon to be Grade 12), from national exams, which are taken by the majority of students. Considering these factors, the OECD review team recommends that Kazakhstan only keep the sample-based Grade 11 EAAA if it serves a clear and unique purpose, complimentary to (not overlapping with) existing tests in the overall student assessment framework. For example, the Grade 11 EAAA could test subjects not covered in other assessments or help develop and pilot test items for use high-stakes examinations (see Recommendation 2.5). Kazakhstan might also consider options that would further reduce the testing burden (and cost) at the end of upper-secondary education, such as keeping the EAAA in Grade 11 when the school system (and examinations) move to Grade 12 and/or administering the test every other year.

Consider developing diagnostic standardised tests for earlier grades

Kazakhstan has no reliable, comparable data at the student level about learning outcomes prior to Grade 4 (when students are around age 9-10). The consolidation of a child's cognitive skills in the early years of

primary school are essential for future learning. For this reason, many OECD countries administer assessments in at least one of the early years of primary school to help identify learning issues before they become problematic and use the data to develop relevant pedagogical responses. For example, Ireland administers its sample-based National Assessment of Mathematics and English Reading (NAMER) during the spring to students in the second year of the primary cycle (second class, ages 7-8) and at the end of the primary cycle (in sixth class, ages 11-12) (ERC, n.d.^[26]). Scotland (United Kingdom), has even designed a full-cohort national assessment for its youngest students (age 4-5) to help secure good outcomes and support their future learning (Scottish Government, n.d.^[27]).

Waiting until the end of primary school to provide teachers with objective data about student learning may be too late if Kazakhstan’s national assessment intends to help teachers, schools and the system identify and address learning gaps. Kazakhstan’s results in TIMSS 2015 show improvements over time in the achievement of Grade 4 students, with the majority reaching the “intermediate benchmark” level in both mathematics (80%) and science (81%) – shares that are higher than the international average and equivalent to the performance of students in many OECD countries (Mullis et al., 2016^[28]). However, having data about individual students and schools, at earlier moments in the education cycle could help develop targeted interventions to further improve learning outcomes.

Several countries have introduced centrally-developed diagnostic assessments in recent years to help improve the quality of teacher-developed assessments and change the perception that low performance signals a problem with individual students rather than a responsibility to support individual learning needs. For example, South Africa recently introduced centrally-developed diagnostic assessments as part of the country’s new integrated assessment framework (see Box 1 in the policy perspective on examinations), which students take in Grade 3 - mid-way through primary school – in addition to Grades 6 and 9.

To further leverage the formative potential of the EAAA, the NTC should consider introducing a diagnostic assessment for all students in an earlier grade of primary education. This assessment should be centrally designed by the NTC to help ensure its reliability and provide teachers with examples of questions they might develop themselves for classroom assessments. However, unlike other large-scale tests developed by the NTC, teachers should be responsible for both administering and marking the diagnostic assessments. This approach would help teachers identify where students might need to review material from the previous year and develop a baseline for evaluating individual student progress. If this diagnostic assessment is to have real formative value, teachers should receive training on how to mark and use the results, in particular how to respond to and manage the diverse learning needs in their classrooms.

2.6. Develop a coherent plan about which subjects the EAAA will assess and align these with national priorities

Evidence

Countries vary in their approach to selecting subjects for national assessments. Some prefer covering a breadth of domains, while others prefer fewer subjects to enable greater depth and content coverage within individual subject areas. At the primary level, around a third of OECD countries with national assessments measure only mathematics and reading and writing in the national language, reflecting a desire to concentrate on foundational skills in the early years of compulsory education (OECD, 2015^[8]). The subjects assessed at the secondary level tend to cover a wider range of subjects, often in areas relevant to the country’s economic development, such as foreign languages, sciences, and civics or social studies. Additional subjects imply additional costs to develop and administer tests; therefore, some countries choose to assess subjects on a rotation or alternate basis.

Kazakhstan lacks strategic direction regarding the subjects that are assessed by the EAAA. Currently, the only mandatory subject is the Kazakh language test in Grade 9. The number of non-mandatory subjects

on the Grade 9 EAAA are determined annually have ranged from three additional subjects (in 2014) to eight subjects (in 2016). Moreover, students participating in the Grade 9 EAAA sample take different tests covering different subject areas, meaning that in a given year, some Grade 9 students might be tested in Algebra and Physics while others are tested in National History, Biology and Computer Science. While this approach aims to cover all areas of the curriculum over time, it reduces the sample size for each domain, undermining the reliability of trend data for the Grade 9 national assessment. There is some consistency in the domains assessed across EAAA cycles and Grades. However, there is no clear timetable or strategy for how often certain subjects should be assessed and in what grades because the ministry and Quality Control Committee have full discretion to select non-mandatory EAAA subjects each year.

The way Kazakhstan currently determines which subjects the EAAA will assess presents a challenge in terms of the cost and capacity needed to develop several valid subject tests that each measure the extent to which students have mastered domains in different grade levels. Intermittent assessment in some domains is an acceptable trade-off to reducing the testing burden on students and schools and limiting costs. However, other domains are more critical to student success and should be assessed regularly. In Kazakhstan, according to the State Programme, mathematics and science are recognised as such domains. However, the EAAA does not regularly assess these subjects, hindering the government's ability to monitor progress in areas important for developing a more innovative and sustainable economy. The lack of consistent trend data on student performance in these areas also prevents the ministry from developing resources to improve subject-specific instruction and adapt the national curriculum, as needed.

Recommended actions

This review recommends that Kazakhstan reflect on the subjects assessed by the EAAA in different grade levels to ensure the instruments serve as valid measures of the national curriculum and complement other large-scale tests administered in the country. Having a national assessment framework (see Recommendation 1 in the policy perspective on examinations) can help the EAAA steering committee plan which subject areas should be covered by the assessment and ensure these align with national priorities.

Make mathematics and literacy mandatory assessment subjects in Grades 4 and 9

In regards to the EAAA in Grades 4 and 9, Kazakhstan already has information about student learning in literacy and mathematics from international assessments that are administered around the same time (TIMSS in Grades 4; PISA in Grade 9/10). However, these assessments are administered once every three to four years and are not aligned to the national curriculum. Making numeracy and literacy mandatory subjects for all students in Grades 4 and 9 would help Kazakhstan better understand and respond to the learning needs of students and emphasise that all children who complete basic education should be supported to achieve national learning goals.

Approach the selection of non-mandatory subjects for assessment systematically

Since annual testing is costly and considering the time it takes to develop valid assessments in multiple subject areas, this review recommends that Kazakhstan more systematically select non-mandatory subjects for the EAAA in Grade 9 (which currently tests an unfixed number of subjects) and Grade 11 (which tests three subjects). This means setting a fixed number of subjects in the Grade 9 EAAA and ensuring that all students are assessed in the same domains. It also means that unlike examinations, for which many countries have developed several elective subject tests to reflect more diverse interests of employers and students (see policy perspective on examinations), the inclusion of subject domains in the national assessment should be primarily relevant to national development goals. For example, the Grade 11 EAAA could test subjects identified by Kazakhstan's education laws and strategic documents as necessary for building a competitive economy and cohesive society, such as English language, science, financial and civic literacy (MoES, 2019_[17]). To this end, Kazakhstan should create a timetable to identify

the frequency that (non-mandatory) subjects will be assessed in both Grades 9 and 11. While this timetable does not need to be public, to prevent teachers from preparing their students for the test, strategically selecting subjects in advance can help the EAAA steering committee anticipate what subject tests need to be developed, increasing their validity as a monitoring tool.

2.7. Prioritise the assessment’s reliability and gradually develop different item types to assess learning wider range of skills

Evidence

The most common types of items that appear on national assessments in OECD countries are multiple-choice responses and closed-format or short answer questions (e.g. providing a numeric solution to a mathematics problem) (OECD, 2013^[6]). These item types are easier and quicker to develop, and they can be marked electronically, removing the need to employ and train individuals as markers (Hamilton and Koretz, 2002^[29]; Anderson and Morgan, 2008^[30]). The latter consideration is important in countries that are aiming to reduce corruption in their assessment systems. Other item types include open-ended writing, performing a task, oral questions and oral presentations. These item types now increasingly appear in national assessments because of their ability to assess broader and more transversal skills. However, they require humans to mark them and time to do so, which can raise concerns about the reliability of students’ results and the extent to which stakeholders trust the assessment to be free of interference (see the policy perspective on examinations for a more detailed discussion of assessment reliability and integrity).

Results from TIMSS and PIRLS reveal that young Kazakhstani students (in Grade 4) perform above the average of their international peers in science, mathematics and literacy (Mullis et al., 2016^[28]). In Grade 8, they continue to be among the top performers in mathematics and sciences, on par with OECD countries like England (United Kingdom), Ireland and the United States. However, when Kazakhstani students take the PISA assessment at age 15, they perform significantly behind OECD countries (OECD, 2019^[3]). This outcome can be explained in part by differences in test design. The TIMSS assessment more closely reflects curriculum facts and processes, while the PISA assessment is competence based and tests functional literacy, mathematics and science skills. For example, PISA often asks students to solve problems set in various real-world contexts, a task many Kazakhstani students seem unprepared to do.

To align national tests more closely with Kazakhstan’s new competence-based curriculum, the ministry is discussing plans to introduce open-ended items to the EAAA. Currently, the EAAA only includes simple multiple-choice questions, which appear to encourage memorisation and lack the capacity to assess higher-order skills (MoES, 2019^[17]). While introducing open-ended items to the EAAA would help assess broader sets of complex competences and higher-order skills, this change risks aggravating legacy concerns about the trustworthiness of Kazakhstan’s national assessments because it will require changes to the way in which tests are marked.

At present, schools in Kazakhstan send completed paper versions of the EAAA to their assigned regional office of the National Testing Centre, which then scans and marks test items electronically. Schools that administer the EAAA via computer send completed tests electronically to their regional NTC office for marking. Results are reported within three days. This process ensures fast and accurate marking, helping to alleviate stakeholders’ concerns about human interference with the assessment results. It is likely the public will perceive changes to EAAA marking procedures, for example, by adding open-ended questions that require high quality assessors and moderation processes (e.g. joint marking, sampled second marking), as undermining the assessment’s statistical reliability.

Recommended actions

Kazakhstan will need to strike a balance between the need to measure learning in ways that do not encourage memorisation and having reliable results that are widely trusted. While literature about the use of closed-format questions to measure higher-order cognitive skills is inconclusive, this review recommends that, in the short term, Kazakhstan prioritise ensuring the EAAA's reliability over introducing more sophisticated, open-format test items. The emphasis should be on improving the EAAA's close-format items to better capture the competence-based curriculum (see Recommendation 3 in policy perspective on examinations). In particular, Kazakhstan should review the difficulty level and placement of distractors in the test's multiple-choice questions and introduce simple constructed-response items (e.g. one-word answers) that can be marked by computers (see Recommendation 2.1 in policy perspective on examinations). Since the Education Modernisation Project with the World Bank plans to improve the technological infrastructure of the NTC's regional offices, this could be an opportunity to invest in any of the technology needed to improve Kazakhstan's automated marking capacity.

In the long-term, Kazakhstan can gradually incorporate more complex open-format questions (e.g. essays or algebraic equations), starting with the EAAA assessment in Grade 11. Not only are students at this age more capable of responding at length to open-format questions but this could also serve as a means to pilot open-ended questions for use in the high-stakes Unified National Test (see the policy perspective about examinations for a more in-depth conversation about open-ended items).

2.8. Review background questionnaires to avoid redundancy and collect information about factors that affect educational equity and instructional quality

Evidence

By linking to students' demographic information, national assessments provide insights into factors that influence learning at the national level and across specific groups. There are several ways to establish such links. The preferred method is by connecting students' results on the national assessment to their data from the national EMIS, which captures many important variables and reduces the need to collect additional data during the assessment. However, many countries also collect non-administrative data through their assessment's background questionnaires. This approach can help capture information related to the practices and perceptions of teachers and schools, in addition to the attitudes of students and parents. Such information can be extremely valuable in understanding different educational contexts and developing pedagogical interventions.

Kazakhstan participates in international assessments and surveys that include background questionnaires for teachers, school principals, students and parents. For example, PISA distributes background questionnaires to the principals of all participating schools and includes a basic questionnaire that all students take, with additional options for collecting information from parents, teachers and more focused areas, such as ICT familiarity and well-being (OECD, 2019^[31]). At the national level, the EAAA currently has a background questionnaire for students and school administrators that collects robust information about factors associated with student performance. For example, the questionnaires ask about pre-school participation, family composition, the availability of home and school resources and students' ambitions for the future (in Grades 9 and 11). These questions are well aligned with international practice; however, the data is not fully exploited (see Recommendation 3) and some of the information could be captured more efficiently through links with Kazakhstan's administrative education data.

At present, the IAC is working with the National Testing Centre to integrate national assessment results with administrative data stored in the NED using unique identifiers for schools and students. For example, it will soon be possible to determine if there are performance disparities on the EAAA between students who benefit from *Vseobuch*, a subsidy provided to socially vulnerable students, and those who do not. This

innovation would offer a reliable proxy for examining socio-economic disadvantage, especially for students in Grade 4 who do not complete a student background questionnaire. Streamlining the collection of background information in this way will also leave more space for the EAAA to ask questions that cannot be sourced from Kazakhstan’s administrative data. Despite these plans, the EAAA database and the NED will be stored separately, meaning that users must manually link the two datasets to conduct analysis. As a result, individuals without strong quantitative skills will likely struggle to use EAAA data to examine relevant questions for policymaking.

Recommended actions

This review recommends that Kazakhstan continue efforts to integrate various education databases, especially linking the EAAA results and the NED. In parallel, the NTC should also review the content of the EAAA background questionnaires and start revising them to avoid duplication with the administrative data that can be provided by the NED once the two datasets are linked. For example, questions about gender, language of instruction (student questionnaire) and other school characteristics, such as location and number of teachers (school administrator questionnaire), should all be accessible through the NED. However, the NTC could rephrase the former question about language of instruction to ask students about what language(s) they speak at home.

Once redundancies in the background questionnaires have been removed, the NTC should consider new questions that can help monitor education goals and inform teaching and learning. For example, since Kazakhstan’s State Programme for Education 2020-25 aims to improve inclusion and equity, the EAAA could gather information about the extent to which students receive tutoring outside of school hours. The administrator questionnaire might ask about the types and quality of professional development available to school leaders or the extent to which they support instruction in their schools by organising professional development opportunities for teachers. Such questions could provide insights for improving teaching and learning, further enhancing the formative value of the EAAA. Adding an optional teacher questionnaire could provide even more contextual information; however, this should not duplicate insights collected from Kazakhstan’s participation in the Teaching and Learning International Survey (TALIS).

Developing background questionnaires is a highly technical process. Questions should aim to collect the information needed, while being as parsimonious as possible. Considering the IAC’s experience in managing and implementing international surveys with sophisticated background questionnaires, such as PISA and TIMSS, the NTC should work with the IAC to review the EAAA’s background questionnaires based on these models.

Recommendation 3. Improve the dissemination and use of national assessment results

This policy perspective has suggested that Kazakhstan refine the purpose and design of its national assessment to collect information that better supports system monitoring and teaching and learning practices. However, collecting information is only the first step to evaluating and improving system performance. The ways in which national assessment findings are disseminated and used are key to supporting national education goals, as are the steps a country takes to avoid misuses of the data and unintended negative consequences.

There are several good practices in terms of how Kazakhstan uses and disseminates EAAA data. For example, the IAC includes analysis of EAAA results in its evaluation and research reports and data is publically available on the National Testing Centre’s website. However, Kazakhstan does not produce a national report of EAAA results, nor does it offer detailed information about the assessment’s technical

aspects. Moreover, results are not analysed, interpreted and disseminated in ways that provide meaningful feedback and information to diverse audiences, especially policymakers and educators. While the previous recommendation examined the EAAA's design, this recommendation focuses on how creating tailored reporting structures can provide stakeholders with relevant insights to inform their work. To ensure the various means of dissemination receive adequate financial resources, the costs associated with their development should be included as part of central planning and be budgeted accordingly.

3.1. Broadly disseminate information about the assessment instrument and the data it produces

Evidence

Disseminating national assessment results gives policymakers and practitioners valuable information they can use to help improve student learning. Disseminating findings also promotes transparency and can apply positive pressure on the system to improve. In addition to reporting results, many countries also publish technical details and assessment procedures, which are important for generating trust in the assessment methodology and therefore the results (Kellaghan, Greaney and Murray, 2009^[16]).

Currently, Kazakhstan does not produce a national report of the EAAA's findings, although the IAC includes some analysis of results in their research and reports. The NTC prepares a press release each year to highlight main findings from the assessment and at the time of this review, EAAA data was also published on the Centre's statistics website, as downloadable PDF and Excel files. While these reporting and dissemination efforts offer aggregate comparisons at the regional and oblast levels and for some contextual factors, such as language of instruction, school location (urban/rural) and gender, they do not fully exploit the data collected by the EAAA background questionnaires. In particular, there is no public reporting of results according to socio-economic background nor distributions of achievement levels in relation to learning standards.

The NTC publishes EAAA data on its website in PDF and Excel format. While it is positive that data is publically available, this approach requires users to rely on their own statistical acumen to draw meaning from the results, leaving many without a good understanding of why the assessment matters and how it can inform their work. There is also no public information about the design, development and implementation of Kazakhstan's national assessment. Until EAAA results and information about the instrument become more accessible, it is unlikely that Kazakhstan's national assessment will serve as a tool for system improvement.

Recommended actions

Develop a national report as the primary source of information about the assessment

This review recommends that Kazakhstan develop a national-level report that summarises and analyses the EAAA's main findings for the entire country. A national report can provide a strong basis for Kazakhstan to communicate information about the EAAA and the evidence it generates about the education system. To be useful, national reports should describe the context and objectives of the assessment, the framework that guides its design, how it is implemented, descriptions of achievement levels and results (Kellaghan, Greaney and Murray, 2009^[16]).

The NTC should design the report, which should present trends over time, disaggregate performance levels and compare achievement across different parts of the country and various sub-groups of the population. Reporting these factors (among many others) represents the minimum level of analysis required to inform policymaking. Such analysis would make use of the valuable data collected by the EAAA background questionnaires, most of which goes unexploited (see Recommendation 2.8). Kazakhstan

should also include both text and illustrative figures in the report, rather than only reproducing the results in tabular format. This approach will help readers draw inferences from the assessment data. Key findings and illustrations from report can later be turned into an infographic, similar to what the IAC developed for the National Report on the State of Education.

Consider publishing a separate technical report

It is important to share detailed information about the assessment's technical aspects, such as the sampling design, scoring techniques, scaling, statistical analyses and quality control (Kellaghan, Greaney and Murray, 2009^[16]). Publishing this information allows researchers and experts to critically evaluate the assessment instrument and provide feedback to the EAAA steering committee to help improve the assessments. Some countries include this type of technical information in their national report (in appendixes), while others produce separate general and technical reports (see Recommendation 4.2 in the policy perspective on examinations). This review recommends that Kazakhstan consider publishing technical information about the EAAA in a separate report to avoid overloading general readers and keeps the general report more accessible.

Create a user-friendly interface to make national assessment data easily accessible

While the national report should serve as the primary source of information about the national assessment and its results, Kazakhstan could further improve dissemination of EAAA data (and foster greater demand for it) by creating a user-friendly interface for the public, researchers and government officials to conduct their own analysis. Similar to the example from Box 2, this interface should allow users to create custom tables and graphics with national assessment data, rather than requiring they search through sets of pre-defined tables located in different PDF and Excel files. The ongoing update to the NTC's website presents an opportune moment to develop and introduce this interface. Importantly, the data of individual students and schools should not be publically available on the platform (see Recommendation 3.5). While publishing student data would create issues for personal privacy, publishing school-level results in Kazakhstan would likely be perceived as attaching stakes to the EAAA and risks worsening equity among schools.

Establish a dedicated webpage for the national assessment

Many countries house national reports and data analysis tools on a central website. This approach allows users to easily access information about the assessment instrument and the results it produces. This review recommends establishing a webpage dedicated to the EAAA as a “one-stop shop” for information about the national assessment. The national and technical reports should populate this page, in addition to the data analysis platform and other materials that may be developed to support the dissemination and use of EAAA results. Kazakhstan should also consider developing a Frequently Asked Questions (FAQ) page to explain key information about the assessment to parents and other users. This practice can help answer commonly asked questions, such as “Why should my child take the EAAA” and “What is the assessment data used for?” Such efforts can help reinforce for formative value of the assessment.

Box 2. Data Explorer tools in the United States

Results from the United States' sample-based National Assessment of Educational Progress (NAEP), are published on the Nation's Report Card website. In addition to general information about the assessment, communication materials, and several thematic reports about the results, the website features a variety of data tools. Notably, the NAEP data explorer, which allows the public to navigate results for all main subject areas assessed, including mathematics, reading, writing, and science and for each Grade covered by the assessment (4, 8 and 11).

Although individual student results are not displayed, the data explorer tool enables users to disaggregate assessment information by various student characteristics, such as race/ethnicity, gender, eligibility to the school lunch program, status as English language learner and disability. Results can also be examined by type of school, geographic location (large city, rural area, etc.) and jurisdiction to compare performance across regions, states and districts. This tool produces customised data tables with different statistics (average scale score, percentage, achievement levels, percentiles and standard deviations), to run statistical tests and regression analysis, create charts and produce tailored reports based on the selected materials. The website also has a NAEP dashboard tool, which allows users to easily visualise achievement gaps according to different variables.

Source: (National Assessment of Educational Progress, n.d.^[32]). The Nation's Report Card, NAEP Data Explorer. <https://www.nationsreportcard.gov/ndecore/xplore/NDE> (accessed 13 November 2020).

3.2. Create tailored reporting structures to communicate results with diverse audiences*Evidence*

Broadly reporting EAAA results to the public is an important part of using the national assessment to promote transparency and inform education policy and teaching practices. However, assessment findings have different implications for different audiences. For example, a reading teacher or curriculum developer might be more interested in knowing which concepts students struggled with on the literacy test, whereas regional or local education officials are likely to be more interested in knowing how the performance of their schools compare with other parts of the country. Communicating national assessment data in more targeted ways can help raise the instrument's value and relevance to support the work of different actors.

Kazakhstan does not currently provide tailored reports of EAAA results. Territorial authorities receive a list of schools that participated in the assessment sample with aggregate raw scores by subject and grade level. Participating schools receive a list of students who took the assessment and their scores across each subject. However, lists of aggregate results do not offer the type of information and analysis that can help inform policy making or raise learning outcomes, such as the distribution of student achievement levels in relation to learning standards.

Reflections about how to disseminate results for different audiences will be especially important once Kazakhstan establishes the EAAA as a full-cohort assessment since this will generate more micro-level data about learning outcomes. Considering the resource demands related to census assessments and in order to ensure that results can help drive improvements across the system, it is critical to communicate the assessments' findings in appropriate forms for interested parties.

Recommended actions

This review recommends that Kazakhstan develop tailored reporting structures to disseminate national assessment results in ways that are relevant and beneficial to diverse audiences, especially policy makers

and educators. Table 3 outlines some of the tailored reports that Kazakhstan should consider in order to raise the EAAA’s value as tool for system improvement. In particular, the NTC should provide more analysis when reporting EAAA results to schools and territorial authorities. For example, the State of Alaska (United States) provides district authorities, schools and teachers with detailed reports that not only include raw scores but also comparative information, such as how students performed against State standards, distributions of achievement levels and performance according to curriculum categories.

Table 3. Suggested reporting structure for national assessment in Kazakhstan

Level of analysis	Target audience	Description and considerations
Oblast and rayon	Akimats	<ul style="list-style-type: none"> • Summarises performance for oblast or rayon, respectively • Includes number of participating schools and students, percentage of scores in each achievement level, median and mean scores and contextual benchmarks relevant to respective administrative area • Can include information about specific schools (e.g., for internal planning purposes) but this information should not serve as a crude measure of school quality • Results of individual students should not be reported at this level to protect personal data; however, aggregate contextual information about student and school characteristics (e.g. language of instruction, socio-economic background, etc.) could be useful
School	School leadership	<ul style="list-style-type: none"> • Includes number of students tested, percentage in each achievement level; median and mean scale scores • Summarises performance for whole school compared to the rayon, oblast and national average with other contextualised benchmarks • May include more detailed information than what is publically available
Class roster	School leadership; teachers	<ul style="list-style-type: none"> • Provides information about the extent to which each student in a class has achieved national learning standards • May include contextualised comparison groups • May contain information about how students responded to each item and the competencies those items assessed • Contains confidential information about individual students and should only be available to school officials • Class-level results should not affect teacher appraisal processes
Student	Parents, guardians and students	<ul style="list-style-type: none"> • If Kazakhstan establishes the EAAA as a census assessment – as recommended by this review- it will be possible to produce individual reports for all students • Provides an individual student’s score, achievement level and information on the extent to which national learning standards have been met • May include contextualised comparison groups • Can serve as a basis for teachers, parents and guardians to discuss an individual student’s performance and potential need for instructional support • Care should be taken to avoid the perception that the results carry stakes, as a result, this review recommends that Kazakhstan focus on the above reporting types before developing individual student reports and actively work to communicate the formative nature of the assessment

Source: Adapted from the types of reports available in (Alaska Department of Education, 2019^[33]). Alaska Department of Education. 2019 Educator Guide to Assessment Reports. https://education.alaska.gov/tis/Assessments/Peaks/EducatorGuide_Assessments_Reports.pdf (accessed 13 November 2020).

Developing national reports and a range of other materials to disseminate EAAA results to different audiences requires skilled analysts who can maximise the utility of the EAAA data and experiment with how to share the information most effectively. In light of the increasing responsibilities tasked to the NTC, as proposed by this review, Kazakhstan will likely need to make considerable investments in human and financial resources to implement a reporting structure like the one outlined in Table 3. To this end, a medium-term implementation plan with a dedicated budget should be created to oversee the rollout of tailored EAAA reports over the next five to ten years.

The IAC would be well placed to support the NTC in improving the reporting and dissemination of EAAA results because the agency has undertaken several initiatives to make evidence about the education

system more accessible and is familiar with reporting results of large-scale assessments. If the IAC and other stakeholder who stand to benefit from more tailored reporting sit on the EAAA steering committee (see Recommendation 1), they could advise the NTC on the form and format of different EAAA reports, help develop protocols for fair, timely and accurate reporting, monitor the effectiveness of different reports and take steps to improve their value to different users. These actors could also provide the steering committee with input on the EAAA's background questionnaires to ensure they collect relevant and useful contextual data (see Recommendation 2.8). Together such efforts stand to help Kazakhstan make better use of the results from other national and international assessments.

3.3. Create opportunities to mobilise assessment data for policy making

Evidence

Disseminating results in ways that are more accessible and relevant to policy makers is an important step towards using assessment data to inform education policies and improve system performance. However, policy making is not a linear process and several factors influence decisions, such as resource availability, public interests and personal views (Kellaghan, Greaney and Murray, 2009^[16]). To translate assessment data into policy action, countries cannot rely on national reports alone but must also create opportunities for policy makers and other stakeholders to engage with and discuss results. Despite Kazakhstan's abundance of data on student learning outcomes, from both national and international assessments, the country lacks systematic processes for mobilising assessment results to drive education reforms.

Recommended actions

In addition to building the NTC's capacity to report and disseminate assessment data, this review recommends that Kazakhstan consider establishing some of the following procedures and practices to integrate the EAAA and other standardised assessments into political negotiations and actions:

- **Involving communities** can help ensure that policy makers understand local issues related to assessments and data they produce. Many countries organise seminars or workshops among policymakers and/or with the public or educators to determine what policy actions to take in response to challenges identified by assessment data. For example, Scotland (United Kingdom) involved teachers, local authority representatives, researchers and policy makers in the development of the government-funded *Assessment is for Learning* programme, which aimed to establish a coherent assessment system focused on learners (Hayward, 2015^[34]). Local and national meetings were also held to bring together thinking from different groups across.
- **Commissioning and considering research** can allow policymakers to interpret assessment findings and decide how best to proceed. For example, the ministry or NTC might conduct or commission researchers to use EAAA data to evaluate the relative influence of private tutoring and other home factors on student performance. In turn, such findings could be published as thematic reports.
- **Reviewing international examples** to provide insights about the policies and practices that other countries have used to address similar challenges. However, countries should always assess whether international policies are appropriate for the national context.

3.4. Develop contextualised benchmarks for reporting information and setting goals

Evidence

A key factor in the usefulness of public dissemination of national assessment results is how contextualised the results are. Context is especially important in large countries like Kazakhstan where student learning

is likely to vary across different parts of the country and among different sub-populations. Collecting adequate contextual information, such as the social and economic circumstances of the community or student, first language of a student (whether this differs from the language of instruction), and any special educational needs, can help countries to monitor equity goals, but also detect and address potential disparities (OECD, 2013^[6]; Kellaghan, Greaney and Murray, 2009^[16]).

Some institutions in Kazakhstan report EAAA data according to language of instruction, gender and geographic location, such as the IAC annual report on the State of Education (IAC, 2019^[4]). Kazakhstan's State Programme for Education also uses the EAAA as a benchmark to set minimum performance goals for students in Grades 4 and 9 (Republic of Kazakhstan, 2019^[9]). While there are indicators about achievement gaps between urban and rural students based on results from PISA and TIMSS, the State Programme 2020-25 has no national benchmarks to measure equity in student achievement according to relevant background information. Moreover, the NTC's previous practice of publishing EAAA results online offered limited contextual information when reporting average scores of individual schools. These practices represent missed opportunities for Kazakhstan to use the EAAA to help improve educational equity.

Recommended actions

Improving the EAAA's background questionnaire, as this review recommends (see Recommendation 2.8) and moving to a census-based assessment, will create more opportunities for Kazakhstan to compare and investigate differences in student learning outcomes. Considering the political and historical context of the country's national assessment system, which emphasises top performers and has served as a punitive accountability measure for schools, this review recommends Kazakhstan develop contextualised national and school-level benchmarks to compare achievement.

While contextualised information about EAAA results should be included in all public reports, identifying specific contextualised benchmarks for different audiences can help the government and schools make meaningful comparisons and set realistic targets for improvement. For example, the Australian Government uses its national assessment to help measure progress towards performance goals of closing gaps in learning outcomes for students with indigenous or low socio-economic backgrounds (see Box 3). Australia has also developed protocols and guidelines for reporting and using data in order to prevent schools and other actors from making over-simplistic comparisons (MCEECDYA, 2009^[35]). This approach can help reduce the emphasis on numerical ranking of schools and make comparisons fairer, since the lack of contextual information often results in schools with the greatest concentration of students from advantaged backgrounds continually being considered the most effective.

The NTC management board should work with the ministry to identify what EAAA benchmarks would be most relevant for the national context. For example, Kazakhstan might compare the results of ungraded or multi-shift schools with each other and set goals to reduce any potential disparities. This approach would take into account differences in the learning environment between in these settings and mainstream schools that have separate classrooms for each grade or operate in single-shifts. Such a benchmark could also help leverage information from the EAAA school administrator questionnaire, which already asks about the number of classrooms in each school. Another comparison Kazakhstan should consider is examining performance differences across schools that share the same language of instruction (i.e. Kazakh with Kazakh schools), rather than only comparing aggregate performance differences between languages (i.e. Kazakh with Russian schools). Regional analysis of the data could take a similar approach by considering performance differences within administrative areas rather than only across them.

Box 3. Australia's contextualised benchmarking of national assessment data

The Australian Government has clear sets of goals related to educational equity. These goals are primarily outlined in the 2008 Melbourne Declaration on Educational Goals for Young Australians, which established long-term directions and aspirations for the country's school system, and the National School Reform Agreement, which set out strategic reforms from 2019 to 2023 to improve student outcomes (see table below). Australia has developed key performance measures to monitor progress towards these goals, including benchmarks on student participation, attainment and achievement in education. The latter is measured using data from the National Assessment Program (NAP) literacy and numeracy tests, which are taken by students in Grades 3, 5, 7 and 9. Since the NAP collects a range of data on student background, including indigenous status, sex, geographical location, disability status, language background and socio-economic background. This disaggregated information helps measure important national benchmarks, such as reducing disparities in student achievement by indigenous background.

Equity goals included in high-level Australian education documents		
	Melbourne Declaration	National School Reform Agreement
Equity-related goals	Australian schooling promotes equity and excellence	Academic achievement improves for all students, including priority equity cohorts (including Aboriginal and Torres Strait Islander students, students living in regional, rural and remote locations, students with a disability and students from educationally disadvantaged backgrounds)
Specific equity-related targets and commitments	'Close the gap' for young Indigenous Australians	Australia considered to be a high quality and high equity schooling system by international standards by 2025
	Provide targeted support to disadvantaged students	At least halve the gap for Aboriginal and Torres Strait Islander students in Year 12 or equivalent attainment rate by 2020, from the 2006 baseline
	Focus on school improvement in low socio-economic communities	

Source: (ACARA, 2019^[36]). Australian Curriculum, Assessment and Reporting Authority. Measurement Framework for Schooling in Australia. <https://www.acara.edu.au/docs/default-source/default-document-library/measurement-framework-for-schooling-in-australia-2019773213404c94637ead88ff00003e0139.pdf?sfvrsn=0> (accessed 13 November 2020).

3.5. Support educators in using assessment results to inform their practice*Evidence*

In addition to supporting system monitoring, national assessments can also support teaching and learning, especially when they are census-based. To do this, countries often create materials and opportunities to help educators better understand the assessment instrument and what implications the results have for their work. Teachers in Alaska (United States), for example, not only receive a class summary report that highlights the strengths and weaknesses of their students in different curriculum areas but also a guidebook dedicated to helping them interpret results (Alaska Department of Education and Early Development, 2019^[37]). Many countries also offer detailed analysis of how students performed on particular test items, which helps identify common errors that teachers should be aware of and try to address. These materials can serve as a basis for schools and teachers to discuss results and develop strategies to address areas of low performance. For example, the French Community of Belgium provides item analysis of its annual

standardised assessments, which examines student performance by topic or learning objective (Enseignement Belgique, n.d.^[38]). This information is accompanied by a commentary on performance differences across educational tracks (e.g. vocational or academic) and used to create pedagogical materials that help teachers of successive cohorts fill the gaps identified by the assessment.

Kazakhstan does not currently produce any materials to enhance the formative value of the EAAA. The lack of support for teachers to understand and use assessment results risks undermining the instrument’s ability to drive educational change. Cycles of assessment reforms in Scotland (United Kingdom) revealed that when teachers lacked support and started to view assessments as an increase to their workload, the policy response was to reduce the volume of assessment, all too often leaving in place easier to manage but less valuable types of assessment (Hayward, 2015^[34]).

Recommended actions

Conduct item analysis

To make the most of Kazakhstan’s national assessment, this review recommends that the NTC develop pedagogical materials to support teachers in using EAAA results to inform their practice. In addition to summary reports for teachers and schools (see Table 3), which can help compare student performance with national, regional and other contextualised benchmarks, Kazakhstan should provide guidelines on how to interpret assessment and conduct item-level analysis of select curriculum areas or constructs. This practice can help educators understand how students performed on each item and the competencies those items assessed. Kazakhstan should also encourage schools and teachers to discuss findings from the EAAA’s results at staff meetings or in other forums.

Inform teacher education programmes

Pre-service and professional development programmes for teachers should consider the assessment framework and sample questions, results and questionnaire responses as valuable sources of information that can inform teaching practices (Kellaghan, Greaney and Murray, 2009^[16]). For example, reviewing sample questions from the (reformed) EAAA during a professional development course would give teachers an opportunity to reflect on how their own assessment practices align with the aims of a competence-based educational approach. Insights from the EAAA can also be used to create pedagogical materials that can be studied during teacher education programmes.

3.6. Give schools relevant benchmarks but avoid public rankings

Evidence

Many census-based national assessments have been used to inform school evaluation and improvement efforts. School-level results can apply internal and external pressure for schools to improve, especially when results are available to the public. Nevertheless, using full-cohort assessments to measure individual schools and teachers can lead to an increased risk of test manipulation and cheating because the assessment is perceived as having high stakes (Greaney and Kellaghan, 2008^[21]). This negative perception of assessment can also adversely impact the quality of teaching, since teachers’ fear of public stigma might lead (for instance) to curriculum narrowing and teaching to the test (OECD, 2013^[6]).

In countries with strong legacies of high-stakes testing, publically available school results, particularly in the form of rankings, can motivate schools to achieve high places in ranking tables instead of focusing on helping students learn. Concerns about rankings are especially prevalent in systems with school choice, since publishing individual school-level results may influence student enrolment (Kellaghan, Greaney and Murray, 2009^[16]). For example, this approach can lead to an oversubscription in schools with top scores or rankings considered as “high quality” and empty seats in schools that do not perform as well.

In Kazakhstan, there will be many potential benefits to providing schools - and those responsible for evaluating and supporting schools - more reliable and objective data about student learning. While expanding the EAAA as a full-cohort assessment and implementing the recommendations proposed in this perspective will help raise the assessment's value for schools, producing school-level results also comes with several risks that need to be addressed. In particular, Kazakhstan's previous national assessment system (ISC) served as a direct accountability measure for schools. This legacy seems to have created general mistrust in the EAAA's results, despite the fact that the assessment no longer has punitive consequences for schools and that the ministry has developed procedures to ensure the instrument's integrity (MoES, 2016^[39]). At this stage, publishing school-level EAAA data, even when contextualised, is likely to generate negative behaviours on the part of schools, making it harder to encourage their meaningful engagement with the assessment.

Concerns about the misuses of the data will also influence decisions on extent to which results for individual schools are made public and in what format. Until recently, the NTC website listed the scores and names of individual schools participating in the EAAA, implying that results were a measure of school quality. This type of direct comparison among schools is highly misleading, as school characteristics and student populations vary greatly across Kazakhstan. An ungraded school, for example, serving a disadvantaged, rural population is likely to have lower EAAA results than an urban one that serves more advantaged, urban students. Listing the average score for participating schools makes it easy for external actors and the media to produce decontextualised school rankings, which could lead to competition for student enrolments. Using national assessment results in this way risks worsening the already uneven student distributions in Kazakhstan's urban areas (see the policy perspective on school evaluation).

Recommended actions

In order to manage and reduce the risks of reinstating a census-based national assessment, Kazakhstan should consider the following actions:

Do not make school-level results public

While there are no plans for the EAAA to sanction teachers or schools, there is a high risk that a full-cohort assessment will be perceived as having stakes and be used to create school rankings. To reduce this risk, the OECD recommends that Kazakhstan avoid publishing individual school-level results. In particular, when the NTC finishes updating its website, the performance of individual schools on the EAAA should *not* be restored online. In Finland, for example, school-level national assessment results are meant to develop and steer (European Commission/EACEA/Eurydice, n.d.^[40]). Therefore, while schools participating in the assessment receive feedback on their own results, these are not made public. In the long-term, if Kazakhstan wishes to start publishing school-level results, the NTC might consider disseminating this information without using numerical scores and/or presenting results in a more contextualised manner to allow public users to consider school performance in more equitable terms. For example, Norway makes the school-level results available through an online platform that allows for comparisons to municipal and national averages but prevents schools from being ranked or compared directly (Norwegian Directorate for Education and Training, n.d.^[41]). Kazakhstan's SEDA platform could be a good way to disseminate such information, as the website already considers EAAA results as one of the indicators used to compute the school quality index but does not allow schools to be sorted (IAC, n.d.^[42]).

Support schools and authorised users to interpret results

It is positive that Kazakhstan plans to use EAAA data to inform the risk-based, desk scan of its potential school review process (see policy perspective on school evaluation). This approach will require that authorised users, such as staff within the Quality Control Committee and the new school inspectorate, are able to access and link school-level EAAA results with data from other sources to develop a comprehensive picture of school performance. In large countries like Kazakhstan, developing effective indicators for this type of desk scans is especially important to help identify what schools are most in need of support and direct resources in a timely manner.

Despite the progress being made to improve the use of EAAA results in school external evaluations, the current practice of providing lists of raw scores for students (to schools) and school averages (to Oblasts and Rayons) does not sufficiently support these actors in interpreting and making use of the results. This review therefore recommends that the NTC provide individual schools and authorised users with more contextual information when it shares school-level results to support school improvement. In particular, these actors should receive information comparing the performance of students in a particular school to national learning standards (rather than just the national average score), and compared to schools operating in similar contexts (see Table 3).

References

- ACARA (2019), *Measurement Framework for Schooling in Australia*, [3
6]
<https://www.acara.edu.au/docs/default-source/default-document-library/measurement-framework-for-schooling-in-australia-2019773213404c94637ead88ff00003e0139.pdf?sfvrsn=0>
(accessed on 30 July 2018).
- Alaska Department of Education (2019), *2019 Educator Guide to Assessment Reports for PEAKS English Language Arts and Mathematics and Alaska Science Assessment*, [3
3]
<http://education.alaska.gov>. (accessed on 13 November 2020).
- Alaska Department of Education and Early Development (2019), *2019 Educator Guide to Assessment Reports for PEAKS English Language Arts and Mathematics and Alaska Science Assessment*, [3
7]
https://education.alaska.gov/tls/Assessments/Peaks/EducatorGuide_Assessments_Reports.pdf
(accessed on 13 March 2020).
- Anderson, P. and G. Morgan (2008), *Developing Tests and Questionnaires for a National Assessment of Educational Achievement*, The World Bank, Washington, D.C. [3
0]
- Bridges, D. (2014), *Education Reform and Internationalisation: The Case of School Reform in Kazakhstan*, Cambridge University Press, Cambridge. [1
3]
- Bürger, S., U. Kroehne and F. Goldhammer (2016), “The transition to computer-based testing in large-scale assessments: Investigating (partial) measurement invariance between modes”, *Psychological Test and Assessment Modeling*, Vol. 58/4, pp. 597-616, [2
5]
<https://www.researchgate.net/publication/322330675>.
- Department of Education and Skills (2016), *Action Plan for Education 2016-2019*, Government of Ireland, Dublin, <https://www.education.ie/en/Publications/Corporate-Reports/Strategy-Statement/Department-of-Education-and-Skills-Strategy-Statement-2016-2019.pdf>. [1
2]
- Enseignement Belgique (n.d.), *Non-certification external assessments [Évaluations externes non certificatives]*, The Teaching Portal in the Wallonia-Brussels Federation, [3
8]
<http://www.enseignement.be/index.php?page=25162&navi=2024> (accessed on 5 April 2020).
- ERC (n.d.), *National Assessments of Mathematics and English Reading (NAMER)*, Educational Research Centre, Ireland, <http://www.erc.ie/studies/namer/overview/> (accessed on 20 February 2020). [2
6]
- European Commission/EACEA/Eurydice (n.d.), *Assuring Quality in Education: Policies and Approaches to School Evaluation in Europe*, Publications Office of the European Union., Luxembourg; <http://dx.doi.org/10.2797/678>. [4
0]
- Government of Republic of Kazakhstan (2017), *About the Program – Digital Kazakhstan state program*, Government decree, <https://digital.kz/en/about-the-program/> (accessed on 22 November 2019). [2
4]
- Greaney, V. and T. Kellaghan (2012), *National Assessments of Educational Achievement, Volume 3: Implementing a National Assessment of Educational Achievement*, The World Bank, [1
8]

Washington DC.

- Greaney, V. and T. Kellaghan (2008), *Assessing National Achievement Levels in Education*, The World Bank, Washington, D.C. [2
1]
- Hamilton, L. and D. Koretz (2002), ““Tests and their use in test-based accountability systems””, in L. Hamilton, B. (ed.), *Making Sense of Test-Based Accountability in Education*, RAND Publishing, , Santa Monica, California. [2
9]
- Hayward, L. (2015), “Assessment is learning: the preposition vanishes”, *Assessment in Education: Principles, Policy and Practice*, Vol. 22/1, pp. 27-43, <http://dx.doi.org/10.1080/0969594X.2014.984656>. [3
4]
- IAC (2019), *National Report on the State and Development of the Education System of the Republic of Kazakhstan 2018*, Information Analytic Centre, Nur-Sultan. [4
1]
- IAC (n.d.), *System for education data analysis (SEDA)*, Information Analytic Centre (IAC), <http://seda.iac.kz/about.php> (accessed on 13 December 2019). [4
2]
- INEE (n.d.), *National Institute of Education Evaluation [Instituto Nacional para la Evaluación de la Educación]*, Mexico, <https://www.mejoredu.gob.mx/sobre-el-inee/historia-inee/> (accessed on 20 February 2020). [1
9]
- Kellaghan, T., V. Greaney and T. Murray (2009), *Using the Results of a National Assessment of Educational Achievement National Assessments of Educational Achievement*. [1
6]
- MCEECDYA (2009), *Principles and Protocols for Reporting on Schooling in Australia*, Ministerial Council for Education, Early Childhood Development and Youth Affairs (MCEECDYA), <http://www.educationcouncil.edu.au/site/DefaultSite/filesystem/documents/Reports%20and%20publications/Publications/Measuring%20and%20reporting%20student%20performance/Principles%20and%20protocols%20for%20reporting%20on%20schooling%20in%20Australia.pdf>. [3
5]
- MoES (2019), *Country Background Report: Kazakhstan*, Ministry of Education and Science, Nur-Sultan. [1
7]
- MoES (2016), “Rules for conducting an external assessment of educational achievements”, Ministry of Education and Science, Nur-Sultan, <http://adilet.zan.kz/rus/docs/V1600013287>. [3
9]
- Mullis, I. et al. (2016), *TIMSS 2015 International Results in Mathematics and Science*, Retrieved from Boston College, TIMSS & PIRLS International Study Center, <http://timss2015.org/download-center/> (accessed on 6 April 2020). [2
8]
- National Assessment of Educational Progress (n.d.), *Nation’s Report Card*, <https://www.nationsreportcard.gov/ndecore/xplore/NDE> (accessed on 13 November 2020). [3
2]
- Newton, P. (2007), “Clarifying the purposes of educational assessment”, *Assessment in Education: Principles, Policy & Practice*, Vol. 14/2, pp. 149-170, <http://dx.doi.org/10.1080/09695940701478321>. [2
0]
- Norwegian Directorate for Education and Training (n.d.), *School gate - Primary school - Overview*, <https://skoleporten.udir.no/> (accessed on 1 May 2020). [4
1]
- OECD (2019), *PISA 2018 Results (Volume I): What Students Know and Can Do*, PISA, OECD [3
1]

- Publishing, Paris, <https://dx.doi.org/10.1787/5f07c754-en>.
- OECD (2019), "PISA 2018 Background questionnaires", in *PISA 2018 Assessment and Analytical Framework*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/67e1518f-en>. [3
1]
- OECD (2015), *Education at a Glance 2015: OECD Indicators*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/eag-2015-en>. [8
]
- OECD (2014), *Reviews of National Policies for Education: Secondary Education in Kazakhstan*, Reviews of National Policies for Education, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264205208-en>. [2
3]
- OECD (2013), *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*, OECD Reviews of Evaluation and Assessment in Education, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264190658-en>. [6
]
- OECD (2009), *Measuring Government Activity*, OECD Publishing, Paris, <https://doi.org/10.1787/9789264060784-en>. [1
1]
- OECD/The World Bank (2015), *OECD Reviews of School Resources: Kazakhstan 2015*, OECD Reviews of School Resources, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264245891-en>. [5
]
- Paulo Santiago, A. et al. (2017), *OECD Reviews of School Resources: Chile*, OECD Publishing, Paris. [2
2]
- Republic of Kazakhstan (2019), *State Program for the Development of Education 2020-2025*. [9
]
- Republic of Kazakhstan (2016), "State Program for the Development of Education 2016-19", 205. [1
]
- Schick, A. (2003), "The Performing State: Reflection on an Idea Whose Time Has Come but Whose Implementation Has Not", *OECD Journal on Budgeting*, Vol. 3/2, <http://dx.doi.org/10.1787/budget-v3-art10-en>. [7
]
- Scottish Government (n.d.), *Scottish National Standardised Assessments (SNSA)*, Smarter Scotland, Scottish Government, <https://standardisedassessment.gov.scot/> (accessed on 20 February 2020). [2
7]
- World Bank (2018), *Population ages 0-14 (% of total population) - Kazakhstan, OECD members | Data*, <https://data.worldbank.org/indicator/SP.POP.0014.TO.ZS?locations=KZ-OE> (accessed on 3 March 2020). [2
]
- World Bank (2017), *Kazakhstan - Education Modernization Project*, <http://documents.worldbank.org/curated/en/340301488682834835/pdf/Kazakhstan-PAD-02092017.pdf> (accessed on 8 August 2019). [1
5]
- World Bank (2012), *Kazakhstan: Student Assessment*, SABER Country Report, Systems Approach for Better Education Results. [1
4]
- World Bank (2004), *Ten Steps Ten Steps to a Results-Based Monitoring and Evaluation System to a Results-Based Monitoring and Evaluation System*, <https://www.oecd.org/dac/peerreviews/World%20bank%202004%2010%20Steps%20to%20a%20Results%20Based%20ME%20System.pdf> (accessed on 4 December 2019). [1
0]

Annex A. Table of key national indicators

#	List of key indicators	Kazakhstan	OECD
Background information			
Economy			
1	GDP per capita, PPP (constant 2011 international \$), 2018*	27 738	40 537
2	GDP per capita growth (annual %), 2018*	4.1	2.3
Society			
3	Population growth (annual %), 2018*	1.3	0.6
4	Population aged 14 years or less (%), 2018*	28.5	17.8
Education indicators			
System			
5	Starting age of compulsory education, 2018***	7	5.7
6	Duration of compulsory education (years), 2017***	9	10.9
Students – net enrolment rates			
7	Pre-primary education (ISCED 0), 2017***	54.9	84.4
	Primary education (ISCED 1), 2018***	87.6	95.6
	Secondary education (ISCED 2 and 3), 2018***	89.4	89.4
8	Tertiary education attainment rate (25 to 34 years-old) (ISCED levels 5 to 8), 2015***	50.3	40.9
9	Share of students enrolled in vocational programmes for upper-secondary education (15 to 19 year olds), 2017***	39.7	43.1
Teachers			
10	Mean age of teachers (TALIS 2018)	40.9	44.1
11	Share of female teachers in secondary education	75.5	58.6
12	Ratio of students to teaching staff (2018) Primary education (ISCED 1)***	19.6	15.3
13	Ratio of students to teaching staff (2018) Secondary education (ISCED 2 and 3)***	7.0	13.7
Finance			
14	Total expenditure on education as a percentage of GDP, all levels 2016***	3.0	5.4
15	Total public expenditure on primary education as a percentage of total government expenditure, 2017 for Kazakhstan, 2016 for OECD average***	1.0	3.5
Learning outcomes			
16	Mean students' performance in reading, PISA 2018****	387	487
17	Mean students' performance in science, PISA 2018****	397	489
18	Mean students' performance in mathematics, PISA 2018****	423	489

Source: * The World Bank, World Bank Indicators: Education, <https://data.worldbank.org/topic/education> (accessed on 17 January 2020)

** UIS, UNESCO Institute for Statistics, <http://data.uis.unesco.org> (accessed on 17 January 2020)

*** OECD (2019), TALIS 2018 Results (Volume I): Teachers and School Leaders as Lifelong Learners, TALIS, <https://doi.org/10.1787/1d0bc92a-en>

**** OECD (2019), PISA 2018 Results (Volume I): What Students Know and Can Do, <https://doi.org/10.1787/5f07c754-en>. *

**** OECD (2019), Skills Matter: Additional Results from the Survey of Adult Skills, OECD Skills Studies, <https://doi.org/10.1787/1f029d8f-en>

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document, as well as any data and any map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The statistical data for Israel are supplied by and are under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at <http://www.oecd.org/termsandconditions>.