79938

Kazakhstan

STUDENT ASSESSMENT

SABER Country Report
2012

Key Policy Areas for Student Assessment	Status
1. Classroom Assessment In Kazakhstan, a publically available official document provides guidelines on classroom assessment. While the Ministry of Education and the National Academy of Education provide outlines of what students are expected to learn in different subject areas at different grade levels, the level of performance to be achieved is not clear. There are limited systematic mechanisms to monitor the quality of classroom assessment practices, and classroom assessment information is required to be disseminated only to the school director and deputy director.	Established ●●●○
2. Examinations The Unified National Testing (UNT) program is an examination administered to students in grade 11. The UNT is used as a secondary school graduation examination and a university entrance examination. Some opportunities that prepare examination staff for work on the UNT include annual non-university training courses on educational measurement and evaluation. Teachers are involved in some examination-related tasks, including the development of examination questions and scoring guides. Currently, there are no mechanisms in place to monitor the consequences of the UNT.	Emerging ●●○○
3. National Large-Scale Assessment (NLSA) The External Assessment of Learning Achievements (EALA) is meant to assess the quality of education services and students' mastery of curricula at the end of basic secondary education and higher education in specific subject areas. EALA was administered for the first time in April 2012. The Government of Kazakhstan provides regular funding for core EALA activities. Some opportunities are available to prepare individuals for work on the EALA, including non-university training courses on educational measurement and evaluation, and funding for attending international programs or workshops on educational measurement and evaluation.	Emerging ●●○○
4. International Large-Scale Assessment (ILSA) Kazakhstan participated in TIMSS 2007 and 2011, and PISA 2009, and has plans to participate in PISA 2012, TIMSS 2015, and PIRLS 2016. A team at the National Center for Education Statistics and Assessment is responsible for carrying out ILSA activities in the country. Several ILSA team members have the necessary training and experience to carry out the required ILSA activities effectively. It is not clear whether decisions based on ILSA results have had a positive impact on students' achievement levels.	Established ●●●○

Systems Approach for Better Education Results

THE WORLD BANK

Introduction

Kazakhstan has focused on increasing student learning outcomes by improving the quality of education in the country. An effective student assessment system is an important component to improving education quality and learning outcomes as it provides the necessary information to meet stakeholders' decision-making needs. In order to gain a better understanding of the strengths and weaknesses of its existing assessment system, Kazakhstan decided to benchmark this system using standardized tools developed under The World Bank's Systems Approach for Better Education Results (SABER) program. SABER is an evidence-based program to help countries systematically examine and strengthen the performance of different aspects of their education systems.

What is SABER-Student Assessment?

SABER-Student Assessment is a component of the SABER program that focuses specifically on benchmarking student assessment policies and systems. The goal of SABER-Student Assessment is to promote stronger assessment systems that contribute to improved education quality and learning for all.

National governments and international agencies are increasingly recognizing the key role that assessment of student learning plays in an effective education system. The importance of assessment is linked to its role in:

(i) providing information on levels of student learning and achievement in the system;
(ii) monitoring trends in education quality over time;
(iii) supporting educators and students with real-time information to improve teaching and learning; and
(iv) holding stakeholders accountable for results.

SABER-Student Assessment methodology

The SABER-Student Assessment framework is built on the available evidence base for what an effective assessment system looks like. The framework provides guidance on how countries can build more effective student assessment systems. The framework is structured around two main dimensions of assessment systems: the types/purposes of assessment activities and the quality of those activities.

Assessment types and purposes

Assessment systems tend to be comprised of three main types of assessment activities, each of which serves a different purpose and addresses different information needs. These three main types are: classroom assessment, examinations, and large-scale, system level assessments.

Classroom assessment provides real-time information to support ongoing teaching and learning in individual classrooms. Classroom assessments use a variety of formats, including observation, questioning, and paper-and-pencil tests, to evaluate student learning, generally on a daily basis.

Examinations provide a basis for selecting or certifying students as they move from one level of the education system to the next (or into the workforce). All eligible students are tested on an annual basis (or more often if the system allows for repeat testing). Examinations cover the main subject areas in the curriculum and usually involve essays and multiple-choice questions.

Large-scale, system-level assessments provide feedback on the overall performance of the education system at particular grades or age levels. These assessments typically cover a few subjects on a regular basis (such as every 3 to 5 years), are often sample based, and use multiple-choice and short-answer formats. They may be national or international in scope.

Appendix 1 summarizes the key features of these main types of assessment activities.

Quality drivers of an assessment system

The key considerations when evaluating a student assessment system are the individual and combined quality of assessment activities in terms of the adequacy of the information generated to support decision making. There are three main drivers of information quality in an assessment system: enabling context, system alignment, and assessment quality.

Enabling context refers to the broader context in which the assessment activity takes place and the extent to which that context is conducive to, or supportive of, the assessment. It covers such issues as the legislative or policy framework for assessment activities; institutional and organizational structures for designing, carrying out, or using results from the assessment; the availability of sufficient and stable sources of funding; and the presence of trained assessment staff.

System alignment refers to the extent to which the assessment is aligned with the rest of the education system. This includes the degree of congruence between assessment activities and system learning goals, standards, curriculum, and pre- and in-service teacher training.

Assessment quality refers to the psychometric quality of the instruments, processes, and procedures for the assessment activity. It covers such issues as design and implementation of assessment activities, analysis and interpretation of student responses to those activities, and the appropriateness of how assessment results are reported and used.

Crossing the quality drivers with the different assessment types/purposes provides the framework and broad indicator areas shown in Table 1. This framework is a starting point for identifying indicators that can be used to review assessment systems and plan for their improvement.

Table 1: Framework for building an effective assessment system, with indicator areas

	Assessment types/purposes		
	Classroom assessment	Examinations	Large-scale, system-level assessment
Enabling context	Policies Leadership and public engagement Funding Institutional arrangements Human resources		
System alignment	Learning/quality goals Curriculum Pre- and in-service teacher training opportunities		
Assessment quality	Ensuring quality (design, administration, analysis) Ensuring effective uses		

Source: World Bank.

The indicators are identified based on a combination of criteria, including:

- professional standards for assessment;
- empirical research on the characteristics of effective assessment systems, including analysis of the characteristics that differentiate between the assessment systems of low- versus high-performing nations; and
- theory—that is, general consensus among experts that it contributes to effective assessment.

Levels of development

The World Bank has developed a set of standardized questionnaires and rubrics for collecting and evaluating data on the three assessment types and related quality drivers.

The questionnaires are used to collect data on the characteristics of the assessment system in a particular country. The information from the questionnaires is then applied to the rubrics in order to judge the development level of the country's assessment system in different areas.

The basic structure of the rubrics for evaluating data collected using the standardized questionnaires is summarized in Appendix 2. The goal of the rubrics is to provide a country with some sense of the development level of its assessment activities compared to best or recommended practice in each area. For each indicator, the rubric displays four development levels—*Latent*, *Emerging*, *Established*, and *Advanced*.

These levels are artificially constructed categories chosen to represent key stages on the underlying continuum for each indicator. Each level is accompanied by a description of what performance on the indicator looks like at that level.

- *Latent* is the lowest level of performance; it represents absence of, or deviation from, the desired attribute.
- *Emerging* is the next level; it represents partial presence of the attribute.
- *Established* represents the acceptable minimum standard.
- *Advanced* represents the ideal or current best practice.

A summary of the development levels for each assessment type is presented in Appendix 3.

In reality, assessment systems are likely to be at different levels of development in different areas. For example, a system may be *Established* in the area of examinations, but *Emerging* in the area of large-scale, system-level assessment, and vice versa. While intuition suggests that it is probably better to be further along in as many areas as possible, the evidence is unclear as to whether it is necessary to be functioning at *Advanced* levels in all areas. Therefore, one might view the *Established* level as a desirable minimum outcome to achieve in all areas, but only aspire beyond that in those areas that most contribute to the national vision or priorities for education. In line with these considerations, the ratings generated by the rubrics are not meant to be additive across assessment types (that is, they are not meant to be added to create an overall rating for an assessment system; they are only meant to produce an overall rating for each assessment type). The methodology for assigning development levels is summarized in Appendix 4.

Education in Kazakhstan

Kazakhstan is an upper middle income country that is experiencing high economic growth, with GDP per capita of $11357 (2011), and annual growth of approximately 6 percent. To meet the needs of its expanding economy, Kazakhstan recognizes the need to increase skill levels and learning outcomes through improving the quality of education, among other strategies. Kazakhstan has made substantial progress in increasing primary and secondary school enrollments, achieving near universal access to basic education. However, providing quality education remains a challenge – PISA 2009 results show that 59 percent of students scored below the basic competency level in mathematics, 58 percent scored below the basic competency level in reading, and 55 percent scored below the basic competency level in sciences.

As of 2011, Kazakhstan has reached near universal primary and secondary enrollment rates, although tertiary enrollment rates are lower than those of high-performing education systems. Progression indicators are strong: the primary completion rate, lower secondary graduation ratio, and tertiary graduation ratio are high relative to comparator countries.

The Government of Kazakhstan is committed to investing in human capital as illustrated in the country's *National Strategic Plan 2020* in which human resources development is one of the five pillars. Moreover, the National Strategy for Industrial Innovation Development includes as one of its objectives a tripling of labor productivity between 2000 and 2015. The Government views modernization of the education system as a way to achieve these goals.

Detailed information was collected on Kazakhstan's student assessment system using the SABER-Student Assessment questionnaires and rubrics. It is important to remember that these tools primarily focus on benchmarking a country's policies and arrangements for assessment activities at the system or macro level. Additional data would need to be collected to determine actual, on-the-ground practices in Kazakhstan, particularly by teachers and students in schools. The following sections discuss the findings by each assessment type, accompanied by suggested policy options. These sections present the analysis of information collected using the SABER-Student Assessment questionnaires and rubrics that is also described in "Student Assessment System Policy Note: Kazakhstan 2012." The suggested policy options were determined in collaboration with key local stakeholders based on Kazakhstan's immediate interests and needs. Detailed, completed rubrics for each assessment type in Kazakhstan are provided in Appendix 5.

Classroom Assessment

Level of development

In Kazakhstan, teachers primarily carry out classroom assessment activities to meet external requirements or information needs, to inform their own teaching and their students' learning, and to meet school-level requirements or information needs. Clear guidelines for classroom assessment are outlined by a formal, publically available system-level document, *Model rules for monitoring current performance progress, interim, and final attestation of students at institutions delivering the general education program at the primary, basic secondary, and general secondary education levels*, a 2008 decree that was revised in December 2009.

Some system-wide resources are available to teachers to carry out classroom assessment activities, including resources that outline what students are expected to learn in different subject areas at different grades or age levels. Specifically, the guidelines in the *State compulsory standard for primary, basic secondary, and general secondary education* document and the *Instructional and methodological guidelines on specifics of teaching in various subject areas at a given level of education* explain education process requirements at each level of primary and secondary education, outline learning objectives for a particular subject at a given grade level, specify what students are expected to learn, and offer brief recommendations on the various forms of classroom assessment activities (theme-based written exams, oral quizzes, and testing, among others). However, the recommendations are only limited to enumerating the format of assessment activities without a detailed explanation for use of assessment information or the application of scoring criteria.

In Kazakhstan, there are few system level mechanisms to ensure that teachers develop skills and expertise in classroom assessment. There are no required courses on classroom assessment included in the pre- or in-service training programs offered by the National Institute of Continuing Education for Managers, Academic, and Teaching Staff, or by the regional Teacher Training Institutes of Continuing Education.

Classroom assessment practices in Kazakhstan are of moderate quality. Classroom assessment practices are generally aligned with the pedagogical and curricular frameworks, and usually provide useful feedback to students. Parents are usually well informed about students' grades. However, classroom assessment activities tend to be mainly about recalling information. Data about student performance on classroom assessment activities are mainly used as an administrative or control tool, rather than as a pedagogical or curricular tool. Scoring is usually fair: errors in the scoring of students' work, uneven application of standards for grading students' work, and grade inflation are rarely serious problems. Limited systematic mechanisms exist to monitor the quality of classroom assessment practices. For example, classroom assessment is a required component of a teacher's performance evaluation and of school inspection and teacher supervision.

Kazakhstan also makes efforts to ensure effective uses of classroom assessment. Teachers are required to report individual student performance to the school director and deputy director. Students' classroom assessment results are recorded in the students' own copy books and teacher record books. There are also adequate required uses of classroom assessment to support student learning. Classroom assessment information is not used as an input for external examination results. Required uses of classroom assessment activities to promote and inform student learning include diagnosing student learning issues, providing feedback to students on their learning, informing parents about their child's learning, planning next steps in instruction, and grading students.

Suggested policy options:

1. *Better alignment of classroom assessment with system learning goals.* This can be done by introducing and making available to teachers a variety of system-wide resources for carrying out classroom assessment activities, including an official curriculum or standards document that outlines the level(s) of performance that students are expected to reach in different subject areas at different grade/age levels, textbooks or workbooks that provide support

for classroom assessment, scoring criteria or rubrics for students' work, and online assessment resources.

2. *Ensure that teachers develop skills and expertise in classroom assessment.* This can be done by introducing varied system-level mechanisms such as pre- and in-service teacher training on classroom assessment, providing a required course on classroom assessment in all teacher training programs, and making available opportunities for teachers to participate in conferences and workshops.

3. *Ensure the quality of classroom assessment by putting in place varied and systematic mechanisms to monitor the quality of classroom assessment practices.* These practices could include introducing an external moderation system that reviews the difficulty of classroom assessment activities and the appropriateness of scoring criteria. Additionally, system-wide reviews of the quality of education that include a focus on classroom assessment can be conducted, and funding for research on the quality of classroom assessment activities and how to improve classroom assessment can be provided as well.

4. *Ensure the effective uses of classroom assessment information.* This could be done by requiring classroom assessment information to be disseminated to all key stakeholders, including students, parents, and education officials, as well as by using classroom assessment information as an input to external examinations.

Examinations

Level of development

Emerging

Kazakhstan's national examination, the Unified National Testing (UNT) program, is administered to students in grade 11 to serve as both student certification for school cycle completion and to select students to university or other higher-education institutions. The Committee for Quality Control in Education and Science and the National Testing Center (NTC) under the Ministry of Education and Science of the Republic of Kazakhstan are responsible for implementing the UNT. The examination includes four compulsory subjects (language of instruction, Kazakh language for schools of Russian instruction and Russian language for schools of Kazakh instruction, mathematics, and history of Kazakhstan) and one optional subject (physics, chemistry, biology, geography, world history, Kazakh literature, Russian literature, English language, German language, French language, or arts).

Efforts to reform the UNT are ongoing. In 2011, the Government of Kazakhstan improved UNT test specifications for certain subjects in which students previously scored poorly, and in 2012, efforts were made to improve the quality of some UNT items, such as those focused on assessing logical thinking in mathematics. The cutoff score was also raised, which may increase the number of students failing to reach this cutoff on the exam. Currently, the UNT is paper-based, but in 2013, the Government of Kazakhstan plans to introduce computer-based testing. In 2013, the Government will also pilot test two separate exams for the UNT: a school-leaving examination and a university entrance examination.

Clear policies for the UNT examination are in place and the UNT has been administered since 2004. However, the number and content of subjects tested has changed over time, which makes it difficult to track trends in results. The UNT is authorized by two formal documents: *State Program for Education Development 2011-2020* and *The Rules on Administering the Unified National Testing*. These documents are available to the public and address some key aspects of the

examination, including governance, distribution of power, responsibilities among key entities, and funding sources.

There are independent attempts to improve the examination by stakeholder groups, including the Ministry of Education, National Center for Education Statistics and Assessment, National Center for State Educational Standards and Testing, and the Committee for Quality Control in Education and Science. New organizational arrangements have been introduced to address security, cheating, and efficiency issues. In addition, test development staff from the National Center for Education Statistics and Assessment received technical assistance under the World Bank Joint Economic Research Program (JERP) to improve the quality of UNT items in accordance with international standards. Experts at the Center are also working on new ability tests to measure students' higher-order competencies and skills.

Regular funding is allocated to cover all core examination activities of the UNT: design, administration, data analysis, data reporting, and staff training. This regular funding does not currently cover research and development activities.

The Committee for Quality and Control in Education and Science and the NTC has all of the required facilities to carry out the examination, including computers for all technical staff, secure storage facilities, and adequate communication tools. However, the Committee and the NTC are newly established as of 2004, and not accountable to an external board or agency. In addition, UNT results are not recognized by other countries, only by certification and selection systems in Kazakhstan.

The examination office is inadequately staffed to effectively carry out the UNT. The country offers limited training opportunities related to examination development, including annual non-university training courses on educational measurement and evaluation, and funds attendance at similar international programs. Although the Ministry of Education and subnational governments ensure that each UNT testing site is adequately staffed, effective administration has been a challenge due to poor training of the test administrators, unclear instructions and guidelines, errors in scoring, weaknesses in test design, and frequent errors in the examination questions.

The UNT is aligned with learning goals and students' opportunities to learn to a certain degree. There is a clear understanding of what the UNT measures. Information on the format of the UNT examinations and sample tests are published on the Committee for Quality Control in Education and Science's website. Study guides that include sample test items and preparation advice are also available for purchase throughout the country. Some students may have greater access than others to these preparation materials depending on access to the internet and the availability of funds to purchase paperback study guides.

Some stakeholder groups, including students and parents, question what the examination measures. The most sensitive issue often raised by parents and students, as well as by the media, is the inadequate quality and validity of the test items. Too many items seem to rely on rote memorization. In a recent survey conducted by "Kazakhstan Today," the vast majority of respondents believed that the UNT is not an objective assessment of students' knowledge. [1] Additional reporting by "Kazpravda" also mentions this criticism against the exam. [2] The Government of Kazakhstan recognizes the need to improve the quality of the UNT, and is addressing some of these concerns with forthcoming "ability tests." In 2013 and 2014 the Ministry of Education plans to pilot a two stage UNT examination: the National (school-leaving) Exam, which is intended to test students' functional literacy and logic, and the Complex Exam (university entrance), which is to be an examination for entrance to higher education institutions with various subject tests that are aligned with possible courses of study.

---

[1] Kazakhstan Today (2012). Available online: http://www.kt.kz/?lang=rus&uin=1133168071&chapter=1153551658

[2] Kazpravda (2012). Available online: http://www.kazpravda.kz/c/1316743767

Teachers are provided with opportunities to learn about the UNT. There are up-to-date compulsory courses or workshops on examinations for teachers, and teachers are involved in some examination-related tasks, such as selecting or creating examination questions and scoring guides, and supervising examination procedures.

There are some procedures and policies in place to ensure the quality of the UNT. There is a comprehensive technical report with restricted circulation – the NTC prepares test specifications, which are approved by the Ministry of Education, but these are not made available to the general public. In addition, internal reviews and observers are used to ensure the quality of the examination.

All students can take the examination regardless of background, location, or ability to pay. However, the results are not credible to some stakeholder groups, including parents and students.

Inappropriate behavior surrounding the examination process is high. Documented cases have involved leakage of examination content prior to the examination, impersonation of another individual, unauthorized communication with other candidates by mobile phone or other method, use of prohibited materials, intimidation of examination supervisors, and other problems.

Student names and results are not confidential, as students' UNT scores are published at each UNT administration site, in two national newspapers, and on a government website.

There are very limited options for students who do not perform well on the examination. Based on the UNT results, students with the highest scores receive government scholarships to enter top higher education institutions that require the highest UNT scores; others who pass the cutoff score may choose less competitive schools or fee-based tuition. The rest of the students who do not pass the cutoff score are left to pass the UNT again in another year or enter education institutions abroad, including Russia, China, and Eastern Europe. According to a recent opinion poll taken in Kazakhstan, 53 percent of people believe that the existing UNT system is not effective in allowing fair and transparent entrance to higher education institutions. A more equitable method for selection to higher education institutions was in fact one of the main goals when the Government of Kazakhstan set up the UNT system. Moreover, there are no mechanisms in place to monitor the consequences of the examination. Analysis of results is usually descriptive in nature, and a more thorough review could yield findings on, among other things, the impact on a candidate of failing the examination, and the predictive validity of the exam.

Suggested policy options:

1. *Ensure that the examination program is supported by effective human resources*. It is very important that there is support for the training and professional development of staff in the examination office so that they can carry out examination activities more effectively, and with no issues. This can be done by introducing a range of opportunities that prepare individuals for work on the UNT, including university graduate programs (masters or doctorate level) and university courses (graduate and non-graduate) on educational measurement and evaluation, as well as internships in the examination office.

2. *Ensure the quality of the UNT*. One way to better ensure quality is to make the UNT technical report available to the general public. Other options include introducing more systematic quality control and assurance mechanisms, such as external observers, external certification or audit of the UNT program, and piloting or field testing UNT questions.

3. *Ensure fairness of the UNT by reviewing the procedures for preventing and addressing inappropriate behavior, and putting in place mechanisms to ensure that inappropriate behavior does not take place*. Ensuring that the UNT is a fair and just exam system is a priority of the Government of Kazakhstan. Thus, establishing public monitoring committees and introducing electronic testing instruments and computer programs to conduct analyses of UNT results will assist in preventing the occurrence of inappropriate behavior, academic misconduct, and dishonesty.

National Large-Scale Assessment (NLSA)

Level of development

From 2005 to 2011, the Interim State Control (ISC) assessment was administered annually to all students in grades 4 and 9 across a rotating set of subjects, including mathematics, native language, foreign language, geography, World Knowledge, physics, biology, history of Kazakhstan, and chemistry. The ISC had four objectives: (i) to carry out an assessment of students' learning achievements; (ii) to evaluate the effectiveness of the learning process; (iii) to develop recommendations for improvement of state compulsory education standards; and (iv) to carry out a comparative analysis of the quality of educational services provided by institutions.

In 2012, the External Assessment of Learning Achievements (EALA) was introduced to replace the ISC[3] in response to the adoption of the *State Program for Education Development 2011-2020.* The EALA aims to assess the quality of education services and the level of students' mastery of curricula at the end of basic secondary education and higher education. The main purposes of the EALA are to: (i) monitor education quality at the system level; (ii) hold schools and educators accountable; and (iii) inform policy design, evaluation, and decision-making. The first EALA in 2011 assessed 10 percent of students in Grade 9 in four subjects: language of instruction (Kazakh or Russian), history of Kazakhstan, algebra, and chemistry. The majority of students took a paper-and-pencil test, but students in the capital city, Astana, and in the Atyrau region took a computer-based test.

The EALA will be adjusted as the Government of Kazakhstan begins a transition to a 12-year school model. The Government plans to administer the EALA annually in Kazakh language and three additional subjects to all students at the final grades of lower secondary, upper secondary, and university education

Most aspects of setting clear policies for the EALA are strong, even though the EALA has only been administered once. There is a formal and publicly available policy—the *Law of the Republic of Kazakhstan On Education*—that authorizes the EALA and establishes it as a form of independent monitoring of the quality of education. There is also a written plan for implementing the EALA, which provides detailed organizational and implementation arrangements.

There is regular funding allocated to the EALA to cover all core assessment activities, including design, administration, analysis, and reporting. This funding anticipates medium- and long-term planning of EALA programs and staff training, but it does not cover research and development activities.

Although the EALA is a new assessment, strong organizational structures are in place. There are permanent agencies associated with the NLSA – the Ministry of Education's National Testing Center (NTC) develops and administers the EALA, and the Division of External Examinations develops related policies. In addition, the NTC is accountable to a clearly recognized body, the Ministry of Education's Committee for Quality Control in Education and Science.

Kazakhstan also has effective human resources for the EALA. The country offers some opportunities to prepare individuals for work on the NLSA, including non-university training courses on educational measurement and evaluation, and funding for individuals to attend similar international programs. For example, the National Center for Education Statistics and Assessment offers professional training courses in the area of education quality monitoring, management, and measurement.

In Kazakhstan, the EALA is aligned with learning goals, although the degree of acceptance of its results by stakeholder groups remains to be seen. The EALA measures performance against learning standards related to the mastery of basic secondary and general secondary education programs. There are also regular

[3] The EALA is similar to the ISC, but educational institutions are not subject to closure based on the results of the EALA. The ISC was abolished in 2011 by Presidential order.

internal reviews of the EALA to ensure that it measures what it intends to measure. Test items are developed by a team of experts (including qualified secondary teachers, candidates and doctors of the sciences, university professors, and others), and then reviewed by secondary school teachers.

Teachers are provided with some opportunities to learn about the EALA through courses or workshops that are offered on a regular basis. In addition, the NTC regularly holds scientific and practical workshops in the area of test development to provide methodological assistance to item writers and reviewers (mostly secondary school teachers).

The EALA has specific procedures in place to ensure quality, including training proctors and administrators, providing administrators with a standard manual, piloting the administration of the assessment, tracking booklets with serial numbers, and using external observers. There are plans to establish a working group in 2012 to improve the EALA procedures. In addition, the EALA has a comprehensive technical report with restricted circulation, which meets the basic requirements of NLSA documentation.

Related to equity, different assessment options are offered to include all groups of students in the EALA, including students in hard-to-reach areas and students who have a different language of instruction – the test is administered in Kazakh or Russian, depending on the student's language of instruction. The next EALA will be held at 154 sites across the country to ensure that students do not have to travel far to participate.

The EALA was administered for the first time on April 20 and 23, 2012, therefore limited information is available regarding uses of the NLSA results. However, in the future, it is important that the EALA continues to be disseminated in an effective way, and used by at least some stakeholder groups in a way that is consistent with the purposes and technical characteristics of the assessment. There should also be some mechanisms in place to monitor the consequences of the EALA.

Suggested policy options:

1. *Ensure the availability of effective human resources for carrying out EALA activities*. This could be done by, for example, introducing varied systematic mechanisms to prepare individuals to work on EALA. Such mechanisms might include introducing university graduate programs and courses specifically focused on educational measurement and evaluation, and making available internships or short-term employment at the NTC and the National Center for Education Statistics and Assessment.

2. *Ensure the quality of the EALA by introducing a variety of mechanisms specifically for this purpose.* Mechanisms might include appointing internal reviewers or observers, introducing external certification or audit, and training scorers to guarantee high inter-rater reliability.

3. *Ensure that EALA results are communicated and used by all stakeholders.* Often, reports on the findings of NLSA exercises, such as the EALA, can be very technical and not relevant for a broad range of stakeholder groups. Thus, a variety of reports and communication mechanisms should be used to better address the diverse information needs of stakeholders. Examples of reports and communication mechanisms include customized summary reports for policy makers; targeted reports for schools, teachers, curriculum developers, and teacher trainers; as well as press releases and special reports for radio and television. Adequate budget will need to be allocated for EALA report preparation and dissemination.

International Large-Scale Assessment (ILSA)

Level of development

Kazakhstan has set clear policies for ILSA activities. Students in Kazakhstan have participated in three ILSAs in the last 10 years: the Trends in International Mathematics and Science Study (TIMSS) 2007 (grade 4); TIMSS 2011 (grades 4 and 8); and the Program for International Student Assessment (PISA) 2009. The Government of Kazakhstan has also taken concrete steps to participate in three upcoming ISLAs: PISA 2012, TIMSS 2015, and PIRLS 2016. Participation in these ILSAs is addressed by a formal public policy document, *State Program for Education Development of the Republic of Kazakhstan 2011-2020*.

There is regular funding for ILSA exercises, allocated at the discretion of the government. The funding typically covers all core activities of the ILSA, including international participation fees, implementation of the assessment exercise in the country, processing and analyzing assessment data, reporting and disseminating the results in Kazakhstan, and attending relevant international expert meetings. Funding does not cover research and development activities.

The human resources needed for ILSA are partially in place. There is a team and national coordinator to carry out ILSA activities under the Director of the National Center for Education Statistics and Assessment. In addition, regional coordinators oversee the organization and implementation of the assessment in each region. The national coordinator is fluent in the language of the assessment. However, the ILSA office is inadequately staffed to carry out these assessments in a completely effective manner. Some of the team members have the necessary training, but most of them do not have extensive training or experience in the area of international assessments.

Kazakhstan provides some opportunities to learn about ILSAs, but only to the ILSA team members; there are training workshops, but no university or online courses on the topic of international assessments.

The ILSA team has been unable to attend all of the international workshops or meetings for the ILSA exercises in which the country has participated. This is mainly due to a lack of funding.

To date, Kazakhstan has met all of the technical standards required to have its data presented in the main displays of the international report for each ILSA exercise in which it has participated. However, Kazakhstan has not contributed new knowledge on ILSAs to the global evidence base.

Kazakhstan's ILSA results and information are regularly disseminated throughout the country in a number of ways. Results for TIMSS 2007 were included in the 2009 and 2010 versions of the National Report on the Status of Education in Kazakhstan, and the results were also published in a TIMSS national report and communicated through media outlets. In addition, products to provide feedback to schools and educators about the ILSA results were sometimes made available. The National Center for Education Statistics and Assessment usually works with subnational departments of education to disseminate results, who in turn work with schools and educators. There is some media coverage of the results in major newspapers, information agencies, and on government websites. Although PISA 2009 results have motivated a five-year national work plan on functional literacy development, ILSA results are used in a limited way to inform decision making in Kazakhstan. It also is not clear that decisions based on ILSA results have had a positive impact on students' achievement levels.

Suggested policy options:

1. *Ensure sustainability of the ILSA program*. This can be done by including implementation of international assessments as a line item in the regular government budget, and by ensuring that the funding provided is sufficient to cover research and development as well as core ILSA activities.

2. *Provide learning opportunities on ILSA as a professional/technical field*. This can be done by providing opportunities to learn about ILSA for university students studying assessment, or a related area, and for professionals or university

staff interested in assessment. Learning opportunities could include university courses on the topic of international assessments, funding for attending international workshops or training on international assessments, and on-line courses on international assessments.

3. *Ensure effective uses of ILSA results.* This can be done by regularly and widely disseminating the country-specific ILSA results and information in Kazakhstan as well as by making customized products that provide feedback about ILSA results to schools and educators.

Appendix 1: Assessment Types and Their Key Differences

	Classroom	Large-scale assessment Surveys		Examinations	
		National	International	Exit	Entrance
Purpose	To provide immediate feedback to inform classroom instruction	To provide feedback on overall health of the system at particular grade/age level(s), and to monitor trends in learning	To provide feedback on the comparative performance of the education system at particular grade/age level(s)	To certify students as they move from one level of the education system to the next (or into the workforce)	To select students for further educational opportunities
Frequency	Daily	For individual subjects offered on a regular basis (such as every 3-5 years)	For individual subjects offered on a regular basis (such as every 3-5 years)	Annually and more often where the system allows for repeats	Annually and more often where the system allows for repeats
Who is tested?	All students	Sample or census of students at a particular grade or age level(s)	A sample of students at a particular grade or age level(s)	All eligible students	All eligible students
Format	Varies from observation to questioning to paper-and-pencil tests to student performances	Usually multiple choice and short answer	Usually multiple choice and short answer	Usually essay and multiple choice	Usually essay and multiple choice
Coverage of curriculum	All subject areas	Generally confined to a few subjects	Generally confined to one or two subjects	Covers main subject areas	Covers main subject areas
Additional information collected from students?	Yes, as part of the teaching process	Frequently	Yes	Seldom	Seldom
Scoring	Usually informal and simple	Varies from simple to more statistically sophisticated techniques	Usually involves statistically sophisticated techniques	Varies from simple to more statistically sophisticated techniques	Varies from simple to more statistically sophisticated techniques

Appendix 2: Basic Structure of Rubrics for Evaluating Data Collected on a Student Assessment System

Dimension	Development Level				
	LATENT (Absence of, or deviation from, attribute)	**EMERGING (On way to meeting minimum standard)**	**ESTABLISHED (Acceptable minimum standard)**	**ADVANCED (Best practice)**	**Justification**
EC—ENABLING CONTEXT					
EC1—Policies					
EC2—Leadership, public engagement					
EC3—Funding					
EC4—Institutional arrangements					
EC5—Human resources					
SA—SYSTEM ALIGNMENT					
SA1—Learning/quality goals					
SA2—Curriculum					
SA3—Pre-, in-service teacher training					
AQ—ASSESSMENT QUALITY					
AQ1—Ensuring quality (design, administration, analysis)					
AQ2—Ensuring effective uses					

Appendix 3: Summary of the Development Levels for Each Assessment Type

Assessment Type	LATENT	EMERGING	ESTABLISHED	ADVANCED
	Absence of, or deviation from, the attribute	*On way to meeting minimum standard*	*Acceptable minimum standard*	*Best practice*
CLASSROOM ASSESSMENT	There is no system-wide institutional capacity to support and ensure the quality of classroom assessment practices.	There is weak system-wide institutional capacity to support and ensure the quality of classroom assessment practices.	There is sufficient system-wide institutional capacity to support and ensure the quality of classroom assessment practices.	There is strong system-wide institutional capacity to support and ensure the quality of classroom assessment practices.
EXAMINATIONS	There is no standardized examination in place for key decisions.	There is a partially stable standardized examination in place, and a need to develop institutional capacity to run the examination. The examination typically is of poor quality and is perceived as unfair or corrupt.	There is a stable standardized examination in place. There is institutional capacity and some limited mechanisms to monitor it. The examination is of acceptable quality and is perceived as fair for most students and free from corruption.	There is a stable standardized examination in place and institutional capacity and strong mechanisms to monitor it. The examination is of high quality and is perceived as fair and free from corruption.
NATIONAL (OR SYSTEM-LEVEL) LARGE-SCALE ASSESSMENT	There is no NLSA in place.	There is an unstable NLSA in place and a need to develop institutional capacity to run the NLSA. Assessment quality and impact are weak.	There is a stable NLSA in place. There is institutional capacity and some limited mechanisms to monitor it. The NLSA is of moderate quality and its information is disseminated, but not always used in effective ways.	There is a stable NLSA in place and institutional capacity and strong mechanisms to monitor it. The NLSA is of high quality and its information is effectively used to improve education.
INTERNATIONAL LARGE-SCALE ASSESSMENT	There is no history of participation in an ILSA nor plans to participate in one.	Participation in an ILSA has been initiated, but there still is need to develop institutional capacity to carry out the ILSA.	There is more or less stable participation in an ILSA. There is institutional capacity to carry out the ILSA. The information from the ILSA is disseminated, but not always used in effective ways.	There is stable participation in an ILSA and institutional capacity to run the ILSA. The information from the ILSA is effectively used to improve education.

Appendix 4: Methodology for Assigning Development Levels

1. The country team or consultant collects information about the assessment system in the country.

2. Based on the collected information, a level of development and score is assigned to each dimension in the rubrics:

- Latent = 1 score point
- Emerging = 2 score points
- Established = 3 score points
- Advanced = 4 score points

3. The score for each quality driver is computed by aggregating the scores for each of its constituent dimensions. For example:

The quality driver, 'Enabling Context,' in the case of ILSA, has 3 dimensions on which a hypothetical country receives the following scores: Dimension A = 2 points; Dimension B = 2 points; Dimension C = 3 points. The hypothetical country's overall score for this quality driver would be: (2+2+3)/3 = 2.33

4. A preliminary level of development is assigned to each quality driver.

5. The preliminary development level is validated using expert judgment in cooperation with the country team and The World Bank Task Team Leader.

For scores that allow a margin of discretion (i.e., to choose between two levels of development), a final decision has to be made based on expert judgment. For example, the aforementioned hypothetical country has an 'Enabling Context' score of 2.33, corresponding to a preliminary level of development of 'Emerging or Established.' Based on qualitative information not captured in the rubric, along with expert judgment, the country team chooses 'Emerging' as the most appropriate level.

6. Scores for certain key dimensions under 'Enabling Context' (in the case of EXAM, NLSA, and ILSA) and under 'System Alignment' (in the case of CLASS) were set as ceiling scores, i.e., the overall mean score for the particular assessment type cannot be greater than the score for these key dimensions. These key variables include formal policy, regular funding, having a permanent assessment unit, and the quality of assessment practices.

Appendix 5: SABER-Student Assessment Rubrics for Kazakhstan

This appendix provides the completed SABER-Student Assessment rubrics for each type of assessment activity in Kazakhstan. In each row of the rubric, the relevant selection is indicated by a thick border and an asterisk. The selection may include a superscript number that refers to the justification or explanation for the selection (as indicated by a thick border and an asterisk). The explanation or justification text can be located in the "Development level rating justifications" section at the end of each rubric. If a row includes a superscript, but not a thick border and an asterisk, this means that insufficient information was available to determine the relevant selection in the row.

KAZAKHSTAN
Classroom Assessment

ENABLING CONTEXT AND SYSTEM ALIGNMENT

Overall policy and resource framework within which classroom assessment activity takes place in a country or system, and the degree to which classroom assessment activity is coherent with other components of the education system.

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT AND SYSTEM ALIGNMENT 1: *Setting clear guidelines for classroom assessment*			
There is no system-level document that provides guidelines for classroom assessment.	There is an informal system-level document that provides guidelines for classroom assessment.	There is a formal system-level document that provides guidelines for classroom assessment.[1] *	This option does not apply to this dimension.
This option does not apply to this dimension.	This option does not apply to this dimension.	The availability of the document is restricted.	The document is widely available.[2] *
ENABLING CONTEXT AND SYSTEM ALIGNMENT 2: *Aligning classroom assessment with system learning goals*			
There are no system-wide resources for teachers for classroom assessment.	There are scarce system-wide resources for teachers for classroom assessment.	There are some system-wide resources for teachers for classroom assessment.[3] *	There are a variety of system-wide resources available for teachers for classroom assessment.
There is no official curriculum or standards document.	There is an official curriculum or standards document, but it is not clear what students are expected to learn or to what level of performance.	There is an official curriculum or standards document that specifies what students are expected to learn, but the level of performance required is not clear.[4] *	There is an official curriculum or standards document that specifies what students are expected to learn and to what level of performance.
ENABLING CONTEXT AND SYSTEM ALIGNMENT 3: *Having effective human resources to carry out classroom assessment activities*			
There are no system-level mechanisms to ensure that teachers develop skills and expertise in classroom assessment.	This option does not apply to this dimension.	There are some system-level mechanisms to ensure that teachers develop skills and expertise in classroom assessment.[5] *	There are a variety of system-level mechanisms to ensure that teachers develop skills and expertise in classroom assessment.

ASSESSMENT QUALITY

Quality of classroom assessment design, administration, analysis, and use.

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ASSESSMENT QUALITY 1: *Ensuring the quality of classroom assessment*			
Classroom assessment practices suffer from widespread weaknesses or there is no information available on classroom assessment practices.	Classroom assessment practices are known to be weak.	Classroom assessment practices are known to be of moderate quality.[6] *	Classroom assessment practices are known to be generally of high quality.
There are no mechanisms to monitor the quality of classroom assessment practices.	There are ad hoc mechanisms to monitor the quality of classroom assessment practices.	There are limited systematic mechanisms to monitor the quality of classroom assessment practices.[7] *	There are varied and systematic mechanisms in place to monitor the quality of classroom assessment practices.
ASSESSMENT QUALITY 2: *Ensuring effective uses of classroom assessment*			
Classroom assessment information is not required to be disseminated to key stakeholders.	This option does not apply to this dimension.	Classroom assessment information is required to be disseminated to some key stakeholders.[8] *	Classroom assessment information is required to be disseminated to all key stakeholders.
There are no required uses of classroom assessment to support student learning.	There are limited required uses of classroom assessment to support student learning.	There are adequate required uses of classroom assessment to support student learning, excluding its use as an input for external examination results.[9] *	There are adequate required uses of classroom assessment to support student learning, including its use as an input for external examination results.

Classroom Assessment: Development-level rating justifications

1. The *Model rules for monitoring current performance progress, interim and final attestation of students at institutions delivering general education programs at the primary, basic secondary and general secondary education levels* is a formal document authorized by the Ministry of Education and Science of the Republic of Kazakhstan by the Decree of the Minister of Education # 125 dated March 18, 2008, and # 590 with additions and amendments as of December 25, 2009.

2. The document is available online, in public libraries, in in-service teacher training courses, at education institutions, and sub-national departments of education.

3. A document that outlines what students are expected to learn in different subject areas at different grade/age levels is typically available to teachers for their classroom assessment activities. *The State compulsory standard for primary, basic secondary, and general secondary education* document provides guidelines to education process requirements at each level of education (primary, basic secondary, and general secondary). Additionally, the Ministry of Education and National Academy of Education jointly develop "education programs" which provide a more detailed outline of what students are expected to learn in different subject areas at different grade levels. Specialized education programs for talented students and students with special needs are also developed. Education programs, however, do not outline the level(s) of performance that students are expected to reach in different subject areas at different grade/age levels. *The instructional and methodological guidelines on specifics of teaching in various subject areas at a given level of education* outline learning objectives for a particular subject at a given grade level, provide the list of recommended textbooks, specify what students are expected to learn and occasionally provide brief recommendations on the various forms of assessment that may be used in the classroom, such as theme-based written exams, oral quizzes, testing etc. However, recommendations are only limited to enumerating the format of assessment activities without any detailed explanation for their uses, expectations, scoring criteria etc.

4. The document outlines what student at different grade/age levels are expected to learn, but does not specify to what performance level.

5. Opportunities to participate in item development for, or scoring of, large-scale assessments or exams are available at the system level to ensure that teachers develop skills and expertise in classroom assessment.

   The National Institute of Continuing Education for Managers, Academic and Teaching Staff carries out and ensure the quality of continuing education and retraining of staff at the secondary, general and primary professional education, improves professional competences of managers' and teachers' corps, and monitors the needs in further training and retraining of teachers in Kazakhstan.

   16 regional Teacher Training Institutes of Continuing Education provide regular pre- and in-service training programs for teachers, develop textbooks, methodological, information and other documents and promote their practical implementation, study teachers' training needs, carry out expert evaluation of curricula, learning materials etc. However, there is no training that specifically targets classroom assessment and there are no required courses on classroom assessment.

6. Classroom assessment practices are generally aligned with pedagogical and curricular framework. It is common to observe classroom assessment practices to rely mainly on multiple-choice, selection-type questions, to be mainly about recalling information, and to mainly be used as administrative or control tool rather than as pedagogical resource. It is rare to observe errors in the scoring or grading of students' work. Additionally, uneven application of standards for grading students' work is rarely a serious problem. Grade inflation is rarely a serious problem and parents are usually well informed poorly informed about students' grades. Classroom assessment practices usually provide useful feedback to students.

7. Classroom assessment is a required component of a teacher's performance evaluation and of school inspection or teacher supervision.

8. Teachers are required to report to the school director and deputy director on individual student's performance.

9. Required uses of classroom assessment activities to promote and inform student learning include diagnosing student learning issues, providing feedback to students on their learning, informing parents about their child's learning, planning next steps in instruction, and for grading students for internal classroom uses.

KAZAKHSTAN
Examinations

ENABLING CONTEXT

Overall framework of policies, leadership, organizational structures, fiscal and human resources in which assessment activity takes place in a country or system and the extent to which that framework is conducive to, or supportive of, the assessment activity.

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 1: *Setting clear policies*			
No standardized examination has taken place.	The standardized examination has been operating on an irregular basis [1] *	The examination is a stable program that has been operating regularly.	This option does not apply to this dimension
There is no policy document that authorizes the examination.	There is an informal or draft policy document that authorizes the examination.	There is a formal policy document that authorizes the examination. [2] *	This option does not apply to this dimension.
This option does not apply to this dimension.	The policy document is not available to the public	The policy document is available to the public. [3] *	This option does not apply to this dimension.
This option does not apply to this dimension.	This option does not apply to this dimension.	The policy document addresses some key aspects of the examination. [4] *	The policy document addresses all key aspects of the examination.
ENABLING CONTEXT 2: *Having strong leadership*			
All stakeholder groups strongly oppose the examination or are indifferent to it.[5]	Most stakeholder groups oppose the examination.	Most stakeholders groups support the examination.	All stakeholder groups support the examination.
There are no attempts to improve the examination by stakeholder groups.	This option does not apply to this dimension.	There are independent attempts to improve the examination by stakeholder groups. [6] *	There are coordinated attempts to improve the examination by stakeholder groups.
Efforts to improve the examination are not welcomed by the leadership in charge of the examination.[7]	This option does not apply to this dimension.	Efforts to improve the examination are generally welcomed by the leadership in charge of the examination.	This option does not apply to this dimension.

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 3: *Having regular funding*			
There is no funding allocated for the examination.	There is irregular funding allocated for the examination.	There is regular funding allocated for the examination. [8] *	This option does not apply to this dimension.
This option does not apply to this dimension.	Funding covers some core examination activities: design, administration, data processing or reporting.	Funding covers all core examination activities: design, administration, data processing and reporting. [9] *	This option does not apply to this dimension.
This option does not apply to this dimension.	Funding does not cover research and development. [10] *	This option does not apply to this dimension.	Funding covers research and development.
ENABLING CONTEXT 4: *Having strong organizational structures*			
The examination office does not exist or is newly established.	The examination office is newly established. [11] *	The examination office is a stable organization.	This option does not apply to this dimension.
The examination office is not accountable to an external board or agency. [12] *	This option does not apply to this dimension.	The examination office is accountable to an external board or agency.	This option does not apply to this dimension.
Examination results are not recognized by any certification or selection system.	Examination results are recognized by certification or selection system in the country. [13] *	Examination results are recognized by one certification or selection system in another country.	Examination results are recognized by two or more certification or selection system in another country.
The examination office does not have the required facilities to carry out the examination.	The examination office has some of the required facilities to carry out the examination.	The examination office has all of the required facilities to carry out the examination. [14] *	The examination office has state of the art facilities to carry out the examination.

(CONTINUED)

LATENT	EMERGING	ESTABLISHED	ADVANCED
●○○○	●●○○	●●●○	●●●●
ENABLING CONTEXT 5: ***Having effective human resources***			
There is no staff to carry out the examination.	The examination office is inadequately staffed to effectively carry out the examination, issues are pervasive.[15] *	The examination office is adequately staffed to carry out the examination effectively, with minimal issues.	The examination office is adequately staffed to carry out the assessment effectively, with no issues.
The country does not offer opportunities that prepare for work on the examination.	This option does not apply to this dimension.	The country offers some opportunities that prepare for work on the examination.[16] *	The country offers a wide range of opportunities that prepare for work on the examination.

SYSTEM ALIGNMENT

Degree to which the assessment is coherent with other components of the education system.

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
SYSTEM ALIGNMENT 1: *Aligning examinations with learning goals and opportunities to learn*			
It is not clear what the examination measures.	This option does not apply to this dimension.	There is a clear understanding of what the examination measures.[17] *	This option does not apply to this dimension.
What the examination measures is questioned by some stakeholder groups.[18] *	This option does not apply to this dimension.	What is measured by the examination is largely accepted by stakeholder groups.	This option does not apply to this dimension.
Material to prepare for the examination is minimal and it is only accessible to very few students.	There is some material to prepare for the examination that is accessible to some students.	There is comprehensive material to prepare for the examination that is accessible to most students.[19] *	There is comprehensive material to prepare for the examination that is accessible to all students.
SYSTEM ALIGNMENT 2: *Providing teachers with opportunities to learn about the examination*			
There are no courses or workshops on examinations available to teachers.	There are no up-to-date courses or workshops on examinations available to teachers.	There are up-to-date voluntary courses or workshops on examinations available to teachers.	There are up-to-date compulsory courses or workshops on examinations for teachers.[20] *
Teachers are excluded from all examination-related tasks.	Teachers are involved in very few examination-related tasks.	Teachers are involved in some examination-related tasks.[21] *	Teachers are involved in most examination-related tasks.

ASSESSMENT QUALITY

Degree to which the assessment meets quality standards, is fair, and is used in an effective way.

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ASSESSMENT QUALITY 1: *Ensuring quality*			
There is no technical report or other documentation.	There is some documentation on the examination, but it is not in a formal report format.	There is a comprehensive technical report but with restricted circulation.[22] *	There is a comprehensive, high quality technical report available to the general public.
There are no mechanisms in place to ensure the quality of the examination.	This option does not apply to this dimension.	There are limited systematic mechanisms in place to ensure the quality of the examination.[23] *	There are varied and systematic mechanisms in place to ensure the quality of the examination.
ASSESSMENT QUALITY 2: *Ensuring fairness*			
Inappropriate behavior surrounding the examination process is high.[24] *	Inappropriate behavior surrounding the examination process is moderate.	Inappropriate behavior surrounding the examination process is low.	Inappropriate behavior surrounding the examination process is marginal.
The examination results lack credibility for all stakeholder groups.	The examination results are credible for some stakeholder groups.[25] *	The examination results are credible for all stakeholder groups.	This option does not apply to this dimension.
The majority of the students (over 50%) may not take the examination because of language, gender, or other equivalent barriers.	A significant proportion of students (10%-50%) may not take the examination because of language, gender, or other equivalent barriers.	A small proportion of students (less than 10%) may not take the examination because of language, gender, or other equivalent barriers.	All students can take the examination; there are no language, gender or other equivalent barriers.[26] *

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ASSESSMENT QUALITY 3: *Using examination information in a fair way*			
Examination results are not used in a proper way by all stakeholder groups. [27]	Examination results are used by some stakeholder groups in a proper way.	Examination results are used by most stakeholder groups in a proper way.	Examination results are used by all stakeholder groups in a proper way.
Student names and results are public. [28] *	This option does not apply to this dimension.	Students' results are confidential.	This option does not apply to this dimension.
ASSESSMENT QUALITY 4: *Ensuring positive consequences of the examination*			
There are no options for students who do not perform well on the examination, or students must leave the education system.	There are very limited options for students who do not perform well on the examination. [29] *	There are some options for students who do not perform well on the examination.	There is a variety of options for students who do not perform well on the examination.
There are no mechanisms in place to monitor the consequences of the examination. [30] *	This option does not apply to this dimension.	There are some mechanisms in place to monitor the consequences of the examination.	There is a variety of mechanisms in place to monitor the consequences of the examination.

Examinations: Development-level rating justifications

1. The Unified National Testing (UNT) is an examination administered to students in grade 11. The examination is used as a secondary school graduation exam and a university entrance exam. The examination includes compulsory and optional subjects.

   From 2004 – 2007, 3 compulsory subjects (Language of instruction (Kazakh or Russian), Mathematics, History of Kazakhstan) and 1 optional subject were tested.

   From 2008 – 2011, 4 compulsory subjects (Native language (language of instruction), Mathematic s, History of Kazakhstan, Kazakh language for schools with Russian language of instruction and Russian language for schools with Kazakh language of instruction) and one optional subject were tested. Optional subjects include: Physics, Chemistry, Biology, Geography, World History, Kazakh Literature, Russian Literature, English language, German language, French language, and arts disciplines.

   In 2013, in addition to introducing computer-based testing, the Government of RK plans to pilot test the separation of UNT functions into two exams (a school-leaving examination and a university entrance examination), scale up this effort by 2014, and finalize the separation in 2015 with extensive use of computer-based testing. Applicants to higher education institutions will have to sit for additional subject exams in the form of independent national testing.

2. The "State Program for Education Development 2011-2020" formal policy document, authorized by the Decree of the President of the Republic of Kazakhstan № 1118 dated December 7, 2010 (Government of the Republic of Kazakhstan), authorizes the examination.

   The Minister of Education also approved "The Rules on administering the Unified National Testing" (The Rules) policy document. This document is also available to the public.

3. The policy document is publicly available for download at the Ministry of Education and other government agencies' websites.

4. The "State Program for Education Development 2011-2020" document outlines governance, distribution of power, responsibilities among key entities, and states funding sources. "The Rules" document specifies, among other procedures, who can sit for the examination, alignment of the examination with curricula and standards, outlines procedures and consequences of cheating and other forms of inappropriate behavior, work of the Review Board, as well as general procedures for disadvantages/special/students with disabilities.

5. Limited information is available support the selection for this dimension.

6. Independent efforts by different stakeholder groups have been made to improve the examination. The Ministry of Education and it's affiliated agencies (National Center for Education Statistics and Assessment, National Center for State Educational Standards and Testing (National Testing Center), Committee for Quality Control in Education and Science) has made coordinated efforts to improve the examination. In addition, in 2012, new organizational arrangements primarily for addressing issues of security, cheating, and efficiency, will be introduced during administration of the UNT. These arrangements include cell phone signal suppressors, videotaping equipment, decrease in the number of exam rooms at each UNT administration site, decrease in the number of State Committee members who are responsible for organizational arrangements during the exam, informational and outreach campaigns among population. Technical assistance is provided by World Bank experts as part of the Kazakhstan-World Bank Joint Economic Research Program to improve the quality of test items based on international standards. World Bank experts worked with test development staff at the National Center for Education Statistics and Assessment under the Ministry of Education (The Center) which houses two laboratories on test development and expert evaluation to deliver capacity building training workshops on item writing and analysis, peer reviewing, score interpretation and other aspects of test design. In addition and continuation of this capacity building exercise, the experts at the Center are currently working further on developing the so-called “tests of new generation” (ability tests) to measure students’ higher-order competencies and skills.

7. Limited information is available support the selection for this dimension.

8. There is regular funding allocated by the government.

9. Funding covers all core examination activities.

10. Funding does not cover research and development.

11. The Committee and the NTC have been operating since 2004.

12. The Committee and the NTC are accountable to the Ministry of Education and Science of RK, Government of RK, and President’s Administration only. They do not report to external independent bodies.

13. Examination results are officially recognized by certification and selection systems in the country.

14. The Committee and the NTC have computers for all technical staff, secure storage facilities, and adequate communication tools (phone, email, internet, etc.). The buildings are also generally secure and there is generally adequate access to computer servers.

15. The Ministry of Education and sub-national governments ensure that each UNT site is adequately staffed. The general procedure for the number and type of staff required is reflected in the Rules on administering the UNT.

Issues that have been identified with the performance of the human resources that are responsible for the UNT include poor training of test

administrators or about unclear instructions and guidelines in administering the examination, errors in scoring that have led to delays in results being reported, weaknesses in test design, and frequent errors in the examination questions.

Generally, the most sensitive issue often raised by parents and children as well as media is the perceived inadequate quality and validity of the test items.

16. Non-university training courses/workshops on educational measurement and evaluation are available in the country on an annual basis. Additionally, funding for attending international programs, courses, and workshops on educational measurement and evaluation are available in the country as well.

17. The examination measures national school curriculum guidelines and standards.

18. Some stakeholder groups question what the examination measures. Students and parents are generally concerned with the perceived inadequacies of the test items as many items test what is viewed as irrelevant factual knowledge in a particular discipline. The Government of RK recognizes these concerns and revises the item bank on a regular basis. Test development experts at the Center under the Ministry of Education of RK are working on the new tests that would measure students' higher-order competences and skills (so-called "ability tests").

19. The needed material to prepare for the examination is accessible by most students (50% to 90% of students), but certain student subgroups may have greater access than others (e.g., due to language issues, location). Specifically, while sample tests are available for download at the Ministry's website, not all students have free access to internet or financial resources to purchase paperback study guides.

    The material on the examination that is publically available includes examples of the types of questions that are on the examination and a report on the strengths and weaknesses in student performance. Information on the format of the examination and sample tests are published on the Ministry of Education website, Committee for Quality Control in Education and Science. UNT study guides are also available for purchase throughout the country and usually include sample test items in relevant disciplines from the actual tests and preparation advice. Annually, the Center prepares and publishes the National Report on the Status of Education in RK. One section of the report presents brief aggregated results of the UNT, including strengths and weaknesses in student performance. However, the presentation of the results is generally limited to basic descriptive statistics.

20. There are compulsory courses or workshops on the examinations available to teachers that are regularly updated. National Testing Center, based on recommendations of departments of education, holds training workshops for school teachers. These workshops include training by international test development experts.

21. Examination-related tasks that are mainly performed by teachers include selecting or creating examination questions and examination scoring guides, and until 2012, supervising the examination procedures.

    Teachers in relevant disciplines are also members of the Review Board (RB) at each UNT administration site which reviews students' appeals with regard to the test items at the exam. One central Republican Committee to review appeals (RCRA) is established to ensure compliance with uniform requirements,

resolve disputes with regard to the exam questions as well as protect UNT participants' rights for the period of UNT administration. The RCRA reviews the validity of appeal(s) to add scores to a certain examinee(s) submitted by the local RB and makes its' final decision. In addition, the RCRA requests the RB to submit copies of the examinees' answer sheets as well as confirmation of compliance with UNT procedures. The RCRA makes decision based on a majority of votes. In case the number of votes is equal, the vote of the RCRA's chair will serve as the determining one. Local RBs are established at each UNT administration site as well. The RBs accept and review students' appeals regarding the content of the exam questions and technical issues, make proposals to the RCRA regarding adding scores to a certain examinee, and informs the examinee on results of the RCRA's decision. The minutes of the RCRA's and RBs' sittings are reflected in the minutes which are stored at the National Testing Center for one year. Members of the State Committee as well as parents whose children will be sitting for the UNT exam in the particular year cannot be the members of the RCRA or RBs.

22. There is a comprehensive technical report but with restricted circulation. The National Testing Center prepares test specifications which are approved by the Ministry of Education. These specifications are not available for general public.

23. Internal review or observers ensure the quality of the examination.

24. Inappropriate behaviors that diminish the credibility of the examination typically include Leakage of the content of an examination paper or part of a paper prior to the examination, impersonation when an individual other than the registered candidate takes the examination copying from other candidates, using unauthorized materials such as prepared answers and notes, collusion among candidates via mobile phones, passing of paper, or equivalent, intimidation of examination supervisors, markers or officials, and provision of external assistance via the supervisor, mobile phone etc.

25. The results are perceived as credible by some stakeholder groups.

26. All students may take the examination, regardless of background (e.g., gender, ethnic group), location (e.g., urban, rural), ability to pay (e.g., transportation, fees) or the like.

27. Limited information is available support the selection for this dimension.

28. Student names and results are public. Specifically, students' UNT scores are published at each UNT administration site. The list of students who received the state education grants based on the UNT results are published in two major national newspapers "Kazakhstanskaya Pravda" (www.kazpravda.kz) and "Egemen Kazakhstan" (www.egemen.kz). The list is also published on the government website.

29. Students who do not perform well on the examination may opt for less selective schools/universities/tracks. According to the Rules on administering the UNT, each examinee may sit for the exam once. In case of failing the examination, the student will have the option to sit for the final general subject examinations at their secondary schools. The students will take the final subject examinations the same year they failed the UNT exam. These exams are designed and administered by the schools themselves (just like final exams) and the results are scored by the Commission on final examinations (the

Commission) whose members consist of heads of education institutions and teachers in relevant subject areas. The content of examinations is approved by the district departments of education. Upon passing the examinations, the students receive the Diploma of General Secondary Education. These students are only allowed to apply to colleges (who grant post-secondary education) but not HEIs that year. In order to apply to HEIs next year, these students may take special comprehensive exams the following year which are administered by the HEIs and overseen by the Department of Higher and Post-Graduate Education at the Ministry of Education and Science of RK.

30. No systemic mechanisms are currently in place to monitor the consequences of the examination. The Center and the National Testing Center analyzes the results of the UNT. However, the analysis is usually limited to descriptive nature, and no studies on the impact or predictive validity take place.

KAZAKHSTAN

National (or System-Level) Large-Scale Assessment (NLSA)

ENABLING CONTEXT

Overall framework of policies, leadership, organizational structures, fiscal and human resources in which NLSA activity takes place in a country or system and the extent to which that framework is conducive to, or supportive of, the NLSA activity.

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 1: *Setting clear policies for NLSA*			
No NLSA exercise has taken place.	The NLSA has been operating on an irregular basis.[1] *	The NLSA is a stable program that has been operating regularly.	This option does not apply to this dimension.
There is no policy document pertaining to NLSA.	There is an informal or draft policy document that authorizes the NLSA.	There is a formal policy document that authorizes the NLSA.[2] *	This option does not apply to this dimension.
This option does not apply to this dimension.	The policy document is not available to the public.	The policy document is available to the public.[3] *	This option does not apply to this dimension.
There is no plan for NLSA activity.	This option does not apply to this dimension.	There is a general understanding that the NLSA will take place.	There is a written NLSA plan for the coming years.[4] *
ENABLING CONTEXT 2: *Having strong public engagement for NLSA*			
All stakeholder groups strongly oppose the NLSA or are indifferent to it.[5]	Some stakeholder groups oppose the NLSA.	Most stakeholders groups support the NLSA.	All stakeholder groups support the NLSA.

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 3: *Having regular funding for NLSA*			
There is no funding allocated to the NLSA.	There is irregular funding allocated to the NLSA.	There is regular funding allocated to the NLSA. [6] *	This option does not apply to this dimension.
This option does not apply to this dimension.	Funding covers some core NLSA activities: design, administration, analysis and reporting.	Funding covers all core NLSA activities: design, administration, analysis and reporting. [7] *	This option does not apply to this dimension.
This option does not apply to this dimension.	Funding does not cover research and development activities. [8] *	This option does not apply to this dimension.	Funding covers research and development activities.
ENABLING CONTEXT 4: *Having strong organizational structures for NLSA*			
There is no NLSA office, ad hoc unit or team.	The NLSA office is a temporary agency or group of people.	The NLSA office is a permanent agency, institution or unit. [9] *	This option does not apply to this dimension.
This option does not apply to this dimension. [10]	Political considerations regularly hamper technical considerations.	Political considerations sometimes hamper technical considerations.	Political considerations never hamper technical considerations.
This option does not apply to this dimension.	The NLSA office is not accountable to a clearly recognized body.	The NLSA office is accountable to a clearly recognized body. [11] *	This option does not apply to this dimension.

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 5: ***Having effective human resources for NLSA***			
There is no staff allocated for running an NLSA.	The NLSA office is inadequately staffed to effectively carry out the assessment.	The NLSA office is adequately staffed to carry out the NLSA effectively, with minimal issues.	The NLSA office is adequately staffed to carry out the NLSA effectively, with no issues. [12] *
The country does not offer opportunities that prepare individuals for work on NLSA.	This option does not apply to this dimension.	The country offers some opportunities to prepare individuals for work on the NLSA. [13] *	The country offers a wide range of opportunities to prepare individuals for work on the NLSA.

SYSTEM ALIGNMENT

Degree to which the NLSA is coherent with other components of the education system.

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
SYSTEM ALIGNMENT 1: *Aligning the NLSA with learning goals*			
It is not clear if the NLSA is based on curriculum or learning standards.	This option does not apply to this dimension.	The NLSA measures performance against curriculum or learning standards. [14] *	This option does not apply to this dimension.
What the NLSA measures is generally questioned by stakeholder groups. [15]	This option does not apply to this dimension.	What the NLSA measures is questioned by some stakeholder groups.	What the NLSA measures is largely accepted by stakeholder groups.
There are no mechanisms in place to ensure that the NLSA accurately measures what it is supposed to measure.	There are ad hoc reviews of the NLSA to ensure that it measures what it is intended to measure.	There are regular internal reviews of the NLSA to ensure that it measures what it is intended to measure. [16] *	This option does not apply to this dimension.
SYSTEM ALIGNMENT 2: *Providing teachers with opportunities to learn about the NLSA*			
There are no courses or workshops on the NLSA.	There are occasional courses or workshops on the NLSA.	There are some courses or workshops on the NLSA offered on a regular basis. [17] *	There are widely available high quality courses or workshops on the NLSA offered on a regular basis.

ASSESSMENT QUALITY

Degree to which the NLSA meets technical standards, is fair, and is used in an effective way.

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ASSESSMENT QUALITY 1: *Ensuring the quality of the NLSA*			
No options are offered to include all groups of students in the NLSA.	This option does not apply to this dimension.	At least one option is offered to include all groups of students in the NLSA.	Different options are offered to include all groups of students in the NLSA. [18] *
There are no mechanisms in place to ensure the quality of the NLSA.	This option does not apply to this dimension.	There are some mechanisms in place to ensure the quality of the NLSA. [19] *	There are a variety of mechanisms in place to ensure the quality of the NLSA.
There is no technical report or other documentation about the NLSA.	There is some documentation about the technical aspects of the NLSA, but it is not in a formal report format.	There is a comprehensive technical report but with restricted circulation. [20] *	There is a comprehensive, high quality technical report available to the general public.
ASSESSMENT QUALITY 2: *Ensuring effective uses of the NLSA*			
NLSA results are not disseminated. [21]	NLSA results are poorly disseminated.	NLSA results are disseminated in an effective way.	This option does not apply to this dimension.
NLSA information is not used or is used in ways inconsistent with the purposes or the technical characteristics of the assessment. [22]	This option does not apply to this dimension.	NLSA results are used by some stakeholder groups in a way that is consistent with the purposes and technical characteristics of the assessment.	NLSA information is used by all stakeholder groups in a way that is consistent with the purposes and technical characteristics of the assessment.
There are no mechanisms in place to monitor the consequences of the NLSA.[23]	This option does not apply to this dimension.	There are some mechanisms in place to monitor the consequences of the NLSA.	There are a variety of mechanisms in place to monitor the consequences of the NLSA.

National (or System-Level) Large Scale Assessment (NLSA): Development-level rating justifications

1. External Assessment of Learning Achievements (EALA) is aimed at assessing the quality of education services and the level of students' mastery of secondary school curricula at the end of basic secondary education and higher education curricula in relevant disciplines. The main purposes of the EALA are to: monitor education quality at the system level; for school or educator accountability; and for policy design, evaluation, and decision making.

   EALA was administered for the first time on April 20 and 23, 2012. Ten percent of students in 9th grade from selected general secondary schools were assessed in four subjects (Kazakh language for schools with Kazakh language of instruction or Kazakh language for schools with Russian language of instruction, History of Kazakhstan, Algebra, and Chemistry).

   Starting in 2013, the Government will begin a transition to a 12-year school model. This transition is expected to be completed by 2020, at which point students will take EALA at grade 10 for further streaming into upper secondary education, (e.g. into humanities or technical fields), at grade 12 as a school leaving exam (to obtain a general secondary education diploma), and at the final year at the university. All three administrations will cover all students in the selected grade. Students will be tested in the Kazakh language (compulsory) and in three additional subjects that will be made known on the day of the test administration.

2. The *Law of the Republic of Kazakhstan "On Education"* authorized by the Decree of the President of the Republic of Kazakhstan (authorized on July 27, 2007 with the latest additions and amendments dated January 9, 2012) is a formal policy document that authorizes the EALA.

3. The policy document is publicly available on the websites of the Ministry of Education and other government agencies.

4. While there is no single stand-alone published large-scale assessment plan for the coming years or future assessment rounds, the law *"On Education,"* with the latest amendments and the State Program for Education Development 2011-2020, provide for the EALA as a form of independent monitoring of the quality of education. On November 7, 2011, the Ministry of Education approved an *Implementation Plan on organization and implementation of EALA in education institutions in the Republic of Kazakhstan for 2011-2012 (the Plan)* which specifies the activities involved in implementing the EALA and the deadlines by which the activities should be completed along with the responsible agencies. On April 6, 2012 the Ministry of Education also approved an *Instruction on implementation of EALA in education institutions of the Republic of Kazakhstan* which provides for detailed organizational and implementation arrangements for EALA in secondary and higher education institutions. This *Instruction* specifies the goal of EALA, grade levels tested, and the mandatory subject to be tested. The other three additional subjects will be annually selected and approved by the Ministry of Education. The *Plan* and the *Instruction* are publicly available documents.

5. Limited information is available support the selection for this dimension.

6. Regular (continuous and predictable) funding is allocated by the government.

7. Activities covered by the funding allocated for the EALA also include long- or medium-term planning of program milestones and staff training.

8. Funding does not cover research and development activities.

9. The National Testing Center is under the Ministry of Education and is responsible for test development and on-site administration of the EALA. The Division of External Examinations within the Committee for Quality Control in Education and Science under the Ministry of Education ("the Committee") is in charge of the policymaking regarding the national large-scale assessment program on behalf of the Ministry of Education. The National Testing Center reports to the Committee which also oversees that the EALA is implemented in compliance with the rules/implementation instructions.

10. The EALA was administered for the first time on April 20 and 23, 2012. Therefore, limited information is available support the selection for this dimension.

11. The NTC, which is responsible for carrying out the EALA, is accountable to the Committee for Quality Control in Education and Science, which oversees the implementation/administration of the EALA on behalf of the Ministry of Education. Both the NTC and the Committee for Quality Control in Education and Science are under the Ministry of Education.

12. There is an adequate number of full-time staff that is responsible for EALA activities. No issues have been identified with the performance of the human resources responsible for the EALA.

13. Non-university training courses/workshops on educational measurement and evaluation, and funding for attending international programs or courses/workshops on educational measurement and evaluation are available in Kazakhstan on an annual basis.

    The National Center for Education Statistics and Assessment (the Center) offers professional training courses in the area of education quality monitoring, management. Some of the training courses focus on education measurement (“New approaches to test development in the context of international experience”, “Managing the process of education measurement in education system”). The NTC also hosts various training workshops on the topic of educational measurement, often with participation of international experts. NTC also holds training workshops/courses annually in the areas of general theory of educational measurement, format of test items, test content criteria, which are led by Prof. Vadim Avanessov, a well respected Russian test development expert and the editor of the journal “Educational Measurement”. In 2007 the NTC became an IAEA member. The 38th IAEA Conference will be held in Kazakhstan in 2012.

14. The Law “On Education" specifies that the EALA is aimed at assessing the quality of education services and at determining the level of mastery in basic secondary and general secondary education programs of the compulsory education standards.

15. The EALA was administered for the first time on April 20 and 23, 2012. Therefore, limited information is available support the selection for this dimension.

16. The NTC is responsible for test development. The NTC’s website provides that development of test items is carried out by the team of test developers and the Experts Council approved by the Ministry of Education from qualified secondary school teachers, candidates and doctors of sciences, university professors etc. Experts’ review of test items is generally performed by secondary school teachers in specific disciplines.

17. Courses or workshops on the EALA are offered to teachers on a regular basis. Most courses provide teachers with relevant resources that they can use in their classrooms.

    NTC annually holds scientific and practical workshops to provide methodological assistance and improve the level of training of item writers and reviewers. These courses and workshops are mostly targeted at individuals involved in test development, item writing, and education measurement in general. Training workshops are provided largely to secondary school teachers involved in item writing or reviewing.

18. Special plans are made to ensure that the EALA is administered to students in hard-to-reach areas, and the assessment is offered in the language of instruction for almost all student groups.

    The EALA will be held at 154 sites across the country which are the same administration sites used to hold the Unified National Testing. This number of sites ensures that students from remote areas do not have to travel far to sit for the examination/assessment. EALA is administered in both the Kazakh and Russian languages which serve as the languages of instruction for most of the student groups in the country.

19. Mechanisms that ensure the quality of the EALA instrument include: all proctors or administrators are trained according to a protocol; there is a standardized manual for large-scale assessment administrators; a pilot is conducted before the main data collection takes place; all booklets are numbered; and external reviewers or observers are employed.

    A publically available list identifies unpermitted behaviors on test day (e.g. changing seats, use of calculators and mobile phones, cheating, leaving the room etc.). The pilot testing of test items for EALA was carried out by the NTC and the Department of preschool and secondary education of the Ministry of Education in December 2011.

20. Technical specifications which are prepared for the EALA in the form of a report, are not publically available.

21. The EALA was administered for the first time on April 20 and 23, 2012. Therefore, limited information is available support the selection for this dimension.

22. The EALA was administered for the first time on April 20 and 23, 2012. Therefore, limited information is available support the selection for this dimension.

23. Limited information is available support the selection for this dimension.

KAZAKHSTAN
International Large-Scale Assessment (ILSA)

ENABLING CONTEXT

Overall framework of policies, leadership, organizational structures, fiscal and human resources in which ILSA takes place in a country or system and the extent to which that framework is conducive to, or supportive of, ILSA activity.

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 1: *Setting clear policies for ILSA*			
The country/system has not participated in an ILSA in the last 10 years.	This option does not apply to this dimension.	The country/system has participated in at least one ILSA in the last 10 years.	The country/system has participated in two or more ILSA in the last 10 years.[1] *
The country/system has not taken concrete steps to participate in an ILSA in the next 5 years.	This option does not apply to this dimension.	The country/system has taken concrete steps to participate in at least one ILSA in the next 5 years.[2] *	This option does not apply to this dimension.
There is no policy document that addresses participation in ILSA.	There is an informal or draft policy document that addresses participation in ILSA.	There is a formal policy document that addresses participation in ILSA.[3] *	This option does not apply to this dimension.
This option does not apply to this dimension.	The policy document is not available to the public.	The policy document is available to the public.[4] *	This option does not apply to this dimension.
ENABLING CONTEXT 2: *Having regular funding for ILSA*			
There is no funding for participation in ILSA.	There is funding from loans or external donors.	There is regular funding allocated at discretion.[5] *	There is regular funding approved by law, decree or norm.
This option does not apply to this dimension.	Funding covers some core activities of the ILSA.	Funding covers all core activities of the ILSA.[6] *	This option does not apply to this dimension.
Funding does not cover research and development activities.[7] *	This option does not apply to this dimension.	This option does not apply to this dimension.	Funding covers research and development activities.

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 3: ***Having effective human resources for ILSA***			
There is no team or national/system coordinator to carry out the ILSA activities.	There is a team or national/system coordinator to carry out the ILSA activities.	There is a team and national/system coordinator to carry out the ILSA activities. [8] *	This option does not apply to this dimension.
This option does not apply to this dimension.	The national/system coordinator or other designated team member may not be fluent in the language of the assessment.	The national/system coordinator is fluent in the language of the assessment. [9] *	This option does not apply to this dimension.
This option does not apply to this dimension.	The ILSA office is inadequately staffed or trained to carry out the assessment effectively. [10] *	The ILSA office is adequately staffed or trained to carry out the ILSA effectively, with minimal issues.	The ILSA office is adequately staffed and trained to carry out the ILSA effectively, with no issues.

SYSTEM ALIGNMENT

Degree to which the ILSA meets technical quality standards, is fair, and is used in an effective way.

LATENT	EMERGING	ESTABLISHED	ADVANCED
●○○○	●●○○	●●●○	●●●●
SYSTEM ALIGNMENT 1: *Providing opportunities to learn about ILSA*			
The ILSA team has not attended international workshops or meetings.	The ILSA team attended some international workshops or meetings. [11] *	The ILSA team attended all international workshops or meetings.	This option does not apply to this dimension.
The country/system offers no opportunities to learn about ILSA.	This option does not apply to this dimension.	The country/system offers some opportunities to learn about ILSA. [12] *	The country/system offers a wide range of opportunities to learn about ILSA.
This option does not apply to this dimension.	This option does not apply to this dimension.	Opportunities to learn about ILSA are available to the country's/system's ILSA team members only. [13] *	Opportunities to learn about ILSA are available to a wide audience, in addition to the country's/system's ILSA team members.

ASSESSMENT QUALITY

Degree to which the ILSA meets technical quality standards, is fair, and is used in an effective way.

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ASSESSMENT QUALITY 1: *Ensuring the quality of ILSA*			
Data from the ILSA has not been published.	The country/system met sufficient standards to have its data presented beneath the main display of the international report or in an annex.	The country/system met all technical standards required to have its data presented in the main displays of the international report.	The country/system met all technical standards required to have its data presented in the main displays of the international report. [14] *
The country/system has not contributed new knowledge on ILSA. [15] *	This option does not apply to this dimension.	This option does not apply to this dimension.	The country/system has contributed new knowledge on ILSA.
ASSESSMENT QUALITY 2: *Ensuring effective uses of ILSA*			
If any, country/system-specific results and information are not disseminated in the country/system.	Country/system-specific results and information are disseminated irregularly in the country/system.	Country/system-specific results and information are regularly disseminated in the country/system. [16] *	Country/system-specific results and information are regularly and widely disseminated in the country/system.
Products to provide feedback to schools and educators about the ILSA results are not made available.	This option does not apply to this dimension.	Products to provide feedback to schools and educators about the ILSA results are sometimes made available. [17] *	Products to provide feedback to schools and educators about ILSA results are systematically made available.
There is no media coverage of the ILSA results.	There is limited media coverage of the ILSA results.	There is some media coverage of the ILSA results. [18] *	There is wide media coverage of the ILSA results.
If any, country/system-specific results and information from the ILSA are not used to inform decision making in the country/system.	Results from the ILSA are used in a limited way to inform decision making in the country/system. [19] *	Results from the ILSA are used in some ways to inform decision making in the country/system.	Results from the ILSA are used in a variety of ways to inform decision making in the country/system.
It is not clear that decisions based on ILSA results have had a positive impact on students' achievement levels. [20] *	This option does not apply to this dimension.	This option does not apply to this dimension.	Decisions based on the ILSA results have had a positive impact on students' achievement levels.

International Large Scale Assessment (ILSA): Development-level rating justifications

1. Kazakhstan has participated in three international large-scale assessments: TIMSS 2007 (grade 4), TIMSS 2011 (grades 4 and 8), and PISA 2009. This rubric is completed primarily with regard to information from carrying out TIMSS 2011.

2. Kazakhstan has taken concrete steps to participate in PISA 2012, TIMSS 2015, and PIRLS 2016.

3. “State Program for Education Development of the Republic of Kazakhstan 2011-2020" is a formal document, authorized by the Decree of the President of the Republic of Kazakhstan № 1118 (Government of RK), December 7, 2010, that addresses participation in ILSA.

4. The policy document is publicly available at the Ministry of Education and other government websites.

5. Allocation of funding for ILSA activities is provided for by the government budget (divided into various budget programs and sub-programs), where implementation of international assessments is part of a certain budget program.

6. Funding covers international participation fees, implementation of the assessment exercise in the country (e.g., printing booklets, travel to schools), processing and analyzing data collected from implementation of the assessment exercise, reporting and disseminating the assessment results in the country/system, and attendance at international expert meetings for the assessment exercise.

7. Funding does not cover research and development activities.

8. The Director of the National Center for Education Statistics and Assessment under the auspices of the Ministry of Education and Science of RK (the Center, www.quality.edu.kz, www.naric.kz) (Ms. Aigul Kultumanova) is the national/system coordinator responsible for the PISA and TIMSS international assessments in Kazakhstan.

   The team at the Center is responsible for carrying out international assessments in the country. The Center houses 4 institutes, one of which is the Institute for Educational Measurement , which monitors the status and education development trends in Kazakhstan, develops proposals/recommendations on improving mechanisms for education quality management in the country. Implementation of the international assessment in the regions of Kazakhstan is carried out with direct technical assistance of the sub-national departments of education. In each region a Regional Coordinator was assigned who was responsible for organization and implementation of the assessment.

   As reported by the Center’s staff, the Center had recently been renamed as well as had undergone a re-organization which would result in the upcoming change of the Center’s organizational structure.

9. The national coordinator is fluent in the language of the assessment.

10. Some of the team members have the necessary training to carry out the required assessment activities effectively as well as previous experience working on ILSAs. Specifically, the team at the Laboratory for International Studies (the Laboratory) at the Center responsible for carrying out the ILSAs consists of a team of 11 employees, four of which have been working at the Laboratory since 2010. While team members have prior working experience in education, teaching, foreign language specialists, most of them do not have extensive experience/training in the area of international assessments. Except for one individual that has prior experience in UNT exam implementation on a regional level, the rest of the staff do not seem to have any training and/or experience in other types of assessment. During times of overlap of international assessments, there can be a shortage of human resources, in which case technical work, such as data entry and logistical arrangements are outsourced. Data collection/processing, any type of data analysis and interpretation of findings are not outsourced. There have been delays in the printing of the test booklets, particularly during the overlap in the implementation of TIMSS and PISA, as the Center's editorial- and-publishing service which, among others, produces print outs of test booklets, does not have sufficient capacity to accommodate a large volume of printing. Occasionally, the printers failed, causing delays as well.

11. Team members have attended some of the meetings. Most of the training opportunities (because of the limited funding) are limited to 1 or 2 staff members, usually including the Director. The team participated in each National Research Coordinators meeting arranged by the IEA from 2009-2011. Participation at the 2011 NRC meeting was limited to the Director of the Laboratory for International Studies only. In 2010, the Center’s team also attended training seminar on international data management (as it pertained to the upcoming TIMSS-2011 assessment) and the use of the supporting software arranged by the IEA-DPC in Hamburg, Germany. The funding for any staff training opportunities is very limited, and participation of all staff members/or participation in all international meetings related to international assessment is not possible.

12. There are no university/on-line courses on the topic of international assessments offered in the country, i.e. there are no system-level regular learning opportunities related to international assessments in Kazakhstan. However, before upcoming assessment studies, the Center holds training workshops for the Regional Coordinators and other individuals responsible for the technical implementation of the assessment in the regions. The National Testing Center also hosts various training workshops on the topic of educational measurement, often with participation of international experts.

13. Opportunities to learn about international assessments are offered to individuals working directly on the specific international assessment exercise.

14. Kazakhstan met all technical standards required to have its data presented in the main displays of the international report.

15. Kazakhstan has not contributed new knowledge on ILSA.

16. Country-specific results from the international assessment (TIMSS 2007, as TIMSS 2011 results will be disseminated in December 2012) are disseminated in a number of ways. The Center publishes an annual National Report on the Status of Education in the Republic of Kazakhstan (the National Report). This report is generally available for download at the Ministry of Education’s website and the website of the Center. The 2009 and 2010 National Reports include a chapter on TIMSS 2007 assessment results. The current link for the National Report 2009 provided for download at the Ministry’s web-site is not active. The National Report 2010 is available for download at the Ministry’s website. In 2009, the Center published the National Report on TIMSS 2007results. The Center's website is currently under construction based on their recent restructuring from the National Center for Education Quality Assessment into the National Center for Education Statistics and Assessment, therefore the report is not currently available for download. Copies of the TIMSS 2007 National

Reports were shared with departments of education in the regions and local executive bodies. Additionally, TIMSS 2007 results were communicated through a press release and were also covered on television, radio and newspapers.

17. Limited information is available on how ILSA results are fed back to schools and educators. The Center normally works with sub-national departments of education to disseminate results. The departments in turn work with schools and educators on a local level. The Pavlodar oblast, for example, provides information on its department of education website, briefly reporting on the significance of international assessments for Kazakhstan, general aspects of the TIMSS-2011 study, sample information etc. The website also indicates that it shared the sample of TIMSS items with local schools for their familiarization with the format of the examination.

18. International assessment results have been discussed in major newspapers and information agencies in the country and press-releases have been published, among other places, on the Ministry of Education and Government of Kazakhstan websites.

    It may be said that the results of TIMSS 2007 results have been adequately covered by the media. For example, Kazakhstan Today, a major information agency, published two press-releases on the results. Kazakhstan law internet portal, which is also a source of the major news in Kazakhstan, published on Kazakhstan's relatively high results on TIMSS. The Ministry of Education and the Government of RK's websites published press releases which briefly discussed the results of TIMSS 2007.

19. Limited information is available on whether and how ILSA results are used to inform policy-making to improve education quality.

    While there are recommendations on how assessment results can be used to inform division making in Kazakhstan, assessment results (specifically from PISA 2009), have been the focus of discussions on student learning in Kazakhstan. The President of RK, for example, instructed the government to prepare a 5-year National work plan on functional literacy development among school children in Kazakhstan which is currently under preparation by the Ministry of Education and other appointed agencies. This policy initiative is largely based on Kazakhstan's disappointing results PISA 2009. Assessment results have also been used to continue dialogue on student learning in the country. In a press release published on the Ministry of Education's website in February 2012, the Deputy Prime Minister and the Minister of Education mention that Kazakh pupils do not have the skills to apply acquired knowledge to real world situations based on the results of PISA 2009 assessment. The Minister of Education emphasized that one of the main problems lies with the lack of good teachers and the low status of teachers in the country. The existing State Program for Education Development 2011-2020 provides for policy measures targeted at improving teacher education programs and raising the status of teachers in Kazakhstan.

20. The Center is in need of in-house technical expertise to carry out an in-depth analysis of assessment result

Acknowledgements

This report was prepared by The World Bank SABER-Student Assessment team in collaboration with Scherezad Latif, World Bank Senior Education Specialist and Task Team Leader for education projects in Kazakhstan, Chelsea Coffin, World Bank Education Specialist, HDNED, and Tatiana Li, World Bank Education Specialist, ECAEDU.

References

Clarke, M. 2012. "What Matters Most for Student Assessment Systems: A Framework Paper." READ/SABER Working Paper Series. Washington, DC: World Bank.

World Bank. 2013. World Bank Development Indicators: Kazakhstan Country Indicator Data. Washington, DC: World Bank. Data retrieved from http://databank.worldbank.org/data on March 12, 2013.

———. 2012. "Student Assessment System Policy Note: Kazakhstan 2012." Unpublished report. Washington, DC: World Bank.

www.worldbank.org/education/saber

The **Systems Approach for Better Education Results (SABER)** initiative produces comparative data and knowledge on education policies and institutions, with the aim of helping countries systematically strengthen their education systems. SABER evaluates the quality of education policies against evidence-based global standards, using new diagnostic tools and detailed policy data. The SABER country reports give all parties with a stake in educational results—from administrators, teachers, and parents to policymakers and business people—an accessible, objective snapshot showing how well the policies of their country's education system are oriented toward ensuring that all children and youth learn.

This report focuses specifically on policies in the area of student assessment.

READ
Russia Education Aid for Development
When children learn, nations prosper

THE WORLD BANK