

Strengthening national examinations in Kazakhstan to achieve national goals

Introduction

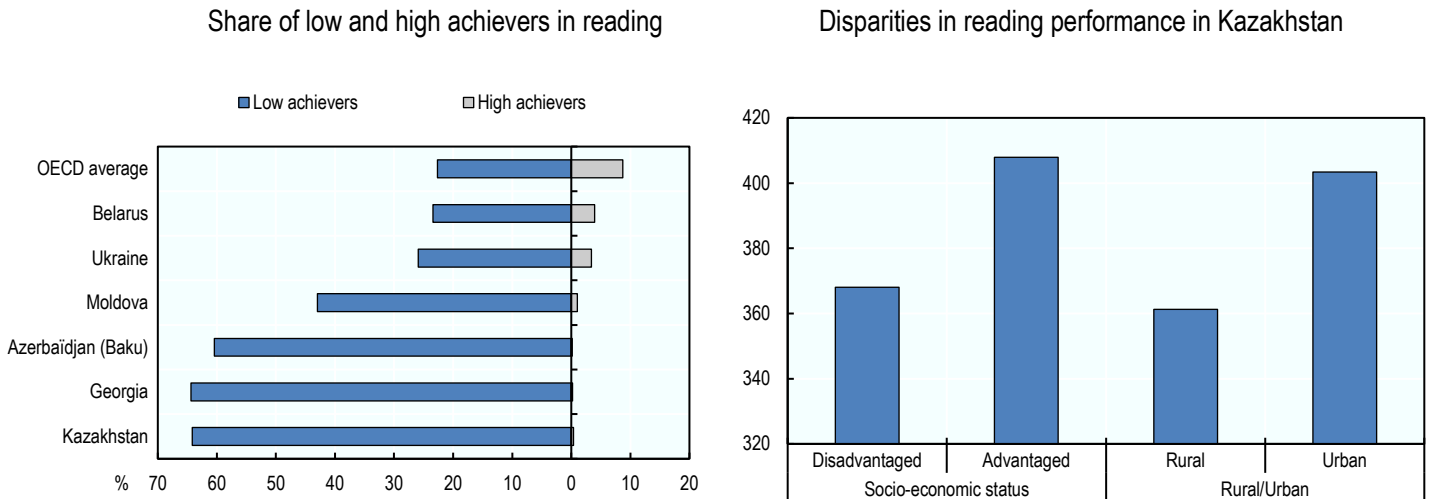
Since achieving independence in 1991, Kazakhstan has undergone rapid development and emerged as a regional economic leader. Recent ambitions, expressed notably in the long-term strategy Kazakhstan 2050, aim to strengthen and diversify the economy in order to position the country as a global leader (Republic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

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

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

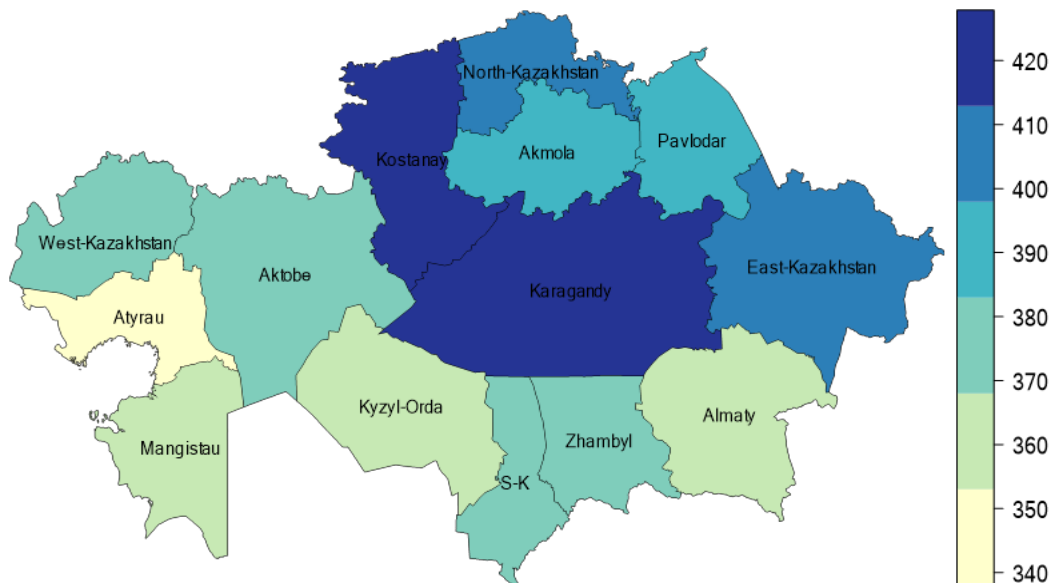
basic minimum standards, influencing their chances of attending tertiary education and finding good employment.

Figure 1. Reading performance in Kazakhstan in PISA 2018



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

Figure 2. Regional differences in reading performance



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

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

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

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

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

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

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

The importance of national examinations

One of the most important tools that education systems use to assess and drive student learning is national examinations. These are tests that are largely standardised at a country-level and carry stakes for students, such as certifying that they have completed a level of education selecting them for entrance into another level. Examinations help achieve national goals by determining how limited educational resources, such as scholarships or places in tertiary institutions, can be efficiently and equitably allocated. They also generate significant educational value by encouraging students, teachers and schools to apply themselves, which helps reinforce the teaching and learning goals of the national curriculum (Bishop, 1999^[6]).

The pivotal position of national examinations also carries several risks. While well-designed systems can help achieve positive outcomes, poorly designed systems can contribute to negative consequences. Examinations that do not assess the most important skills might select students for opportunities for which they are not prepared or in which they are not interested. These examinations might also motivate teachers

to deviate from the curriculum to help students prepare, which can impede progress towards achieving national teaching and learning goals.

Key elements of the education landscape of Kazakhstan

Although this policy perspective focuses on national examinations, it is important to recognise that examinations exist in a larger educational environment. The content of examinations is influenced by national learning standards and the curriculum. Likewise, the extent to which examinations cover the curriculum affects what teachers teach in the classroom. Because examinations both influence and are influenced by the educational landscape as a whole, this policy perspective first discusses the education landscape in Kazakhstan before analysing how the national examinations function in this context.

Learning standards and curriculum

The fundamental role of an education system is to help students develop the knowledge and skills that they need to be successful. To define what students are expected to learn, most countries develop a set of national learning standards that set out what students should know and be able to do at different stages of their education.

In Kazakhstan, the State Compulsory Education Standards, adopted in 2013, are considered the highest level learning standards of the country. The standards were developed by the Y. Altynsarin National Academy of Education (NAE), which is also responsible for developing the “standard curricula” for different subjects and grade levels. Significant changes to the standard curricula in 2016-17 moved Kazakhstan away from fact-based education and towards more competence-based teaching and learning. According to these reforms, it is not only important for students to acquire academic content knowledge, but also to be able to use that knowledge in novel situations to solve problems. Such fundamental changes to learning expectations and the curriculum affect nearly all other parts of the education system (Kitchen et al., 2019^[7]). Kazakhstan’s education system, in particular its examinations, is still adapting to these changes.

Classroom environment

Successfully implementing a competence-based curriculum entails considerable changes in how teachers teach and assess students. For example, determining if a student has mastered a competence requires different assessment methods than determining if a student has memorised a fact. A competence-based approach also implies that a student’s performance is judged in relation to the competences that they are expected to acquire and not in relation to other students’ performance. Changing such practices has been one of the biggest impediments that countries encounter when trying to embed competence-based educational approaches (Baartman et al., 2007^[8]).

Teaching in Kazakhstan is strongly rooted in a traditional understanding that the goal of education is for students to acquire large amounts of knowledge with some neglect for higher-order cognitive skills. This understanding extends to assessment. How well a student is performing is determined by how much knowledge they can demonstrate that they have acquired (which knowledge to acquire was historically determined by individual teachers). In addition to focusing on content knowledge, assessment methods are competitive. Until recently, students’ classroom marks were based upon their performance compared to others in their class (a practice referred to as “norm-referenced assessment”). The same approach encourages students to compete in regional, national and international competitions, with the highest performers (and their teachers and schools) being rewarded for their efforts (OECD, 2014^[9]).

Kazakhstan has introduced several initiatives to make classroom practices better aligned with the aims of a competence-based educational approach. Chief among these is a series of national training programmes

that support teachers in teaching according to the new curriculum (see the policy perspective on initial teacher education), and a law, enacted in 2016-17, that requires teachers to practise criterion-based assessment. According to this law, teachers are to assess students based upon their mastery of competences that are set out in the curriculum, independently of how other students in their class perform (OECD, 2018^[10]). However, these efforts have only been recently introduced and most schooling environments are very different from NIS, where the initiatives were incubated. Therefore, many teachers around the country are still not familiar with how to implement the new curriculum and criterion-based assessment. In such an environment, national examinations serve a critical role because the pressure they exert can influence what teachers teach and how they assess their students.

Choice and certification in upper secondary education

In Kazakhstan, primary education is considered as Grades 1 through 4 and lower secondary education as Grades 5 through 9. At the end of Grade 9, students take an examination to certify completion of lower secondary education and can choose to continue to either general or vocational upper secondary schooling. The specifications for the lower secondary certification examination are developed regionally and, since the test is not standardised, its results are not comparable across regions, which raises issues about the value of the certification it confers. While this policy perspective focuses on examinations at the threshold of upper secondary and tertiary education, there is clearly a need to review the lower secondary certification system. Standardising the examination would be a positive step, such as by centrally developing the items along with the specifications.

In 2018, roughly 60% of Grade 9 graduates entered general upper secondary school while 39% entered vocational education (IAC, 2019^[4]). General upper secondary school ends in Grade 11 (soon to be Grade 12), at which point students must take the Final Attestation¹ to certify completion. Vocational education occurs in colleges in Kazakhstan, rather than schools. Colleges are diverse and offer an array of specialisations, lengths of programmes, and can admit only upper secondary age students or adult learners. College students are certified by either passing a vocational examination or completing a graduation project, both of which are developed and administered at the school-level. Previous OECD reviews have raised concerns with this system, in particular regarding the lack of permeability between the tracks and the lack of opportunity that vocational students have to develop their core academic skills (OECD, 2014^[9]).

The Final Attestation is not required to be taken by students from private upper secondary schools, though these students only represent one percent of all upper secondary students. Some public schools are nationally designated as specialty mathematics schools and students from these schools have an extra task on the algebra and analysis subject of the Final Attestation. Depending upon the elective that students from these schools choose on the Final Attestation, the content of the elective subject might also change. Further, NIS operates according to an internal certification system, and students from these schools do not take the Final Attestation (e.g., many students study an International Baccalaureate curriculum).

Selection and into tertiary education

In order to enrol in university and receive scholarship funding, students, including those from private and specialty schools, must pass the Unified National Test (UNT). An exception is made for NIS graduates, whose marks on their internal examinations are converted into a UNT result. College graduates can also

¹ In Kazakhstan, both the examination after lower secondary education and the examination after upper secondary education are referred to as a “Final Attestation”. For the purposes of this paper, “Final Attestation” only refers to the examination after upper secondary education.

enrol in universities through taking the UNT (students who wish to continue in the same field that they studied in their college take a shorter UNT with fewer subjects).

Data suggest that successfully completing higher education strongly contributes to finding successful employment in Kazakhstan (OECD, 2017_[11]). In 2018, the average salary of a recent graduate with a bachelor degree was over double that of someone with an upper secondary diploma, and the unemployment rate of a higher education degree holder was half that of an upper secondary diploma holder (IAC, 2019_[4]). In fact, returns to education are greater in Kazakhstan than across OECD countries. According to the OECD's Survey of Adult Skills 2017-18, an increase of roughly three years of educational attainment is associated with a 22% increase in hourly wages, compared with less than an 18% increase across the OECD. These differences in outcomes have contributed to strong student demand for more tertiary education. In 2017-18 over 50% of Kazakhstanis between the ages of 25 and 34 held a tertiary degree (higher than the OECD average of 41%), compared to 28% of adults between the ages of 45 to 55.

Contributing to the competition around entering tertiary education is the low public financing of universities. In 2014, Kazakhstan spent 0.3% of its GDP on tertiary education, which is far below the OECD average of 1.6%. In 2017, tuition at public institutions reached up 2 000 USD (at 2016 conversion rates), which is out of reach for many families in Kazakhstan (OECD, 2017_[11]). To assist students, the government offers a limited number of public scholarships that are allocated based on field of study and a student's result on the UNT. In 2018, around 62% of students who applied for a scholarship received one (IAC, 2019_[4]).

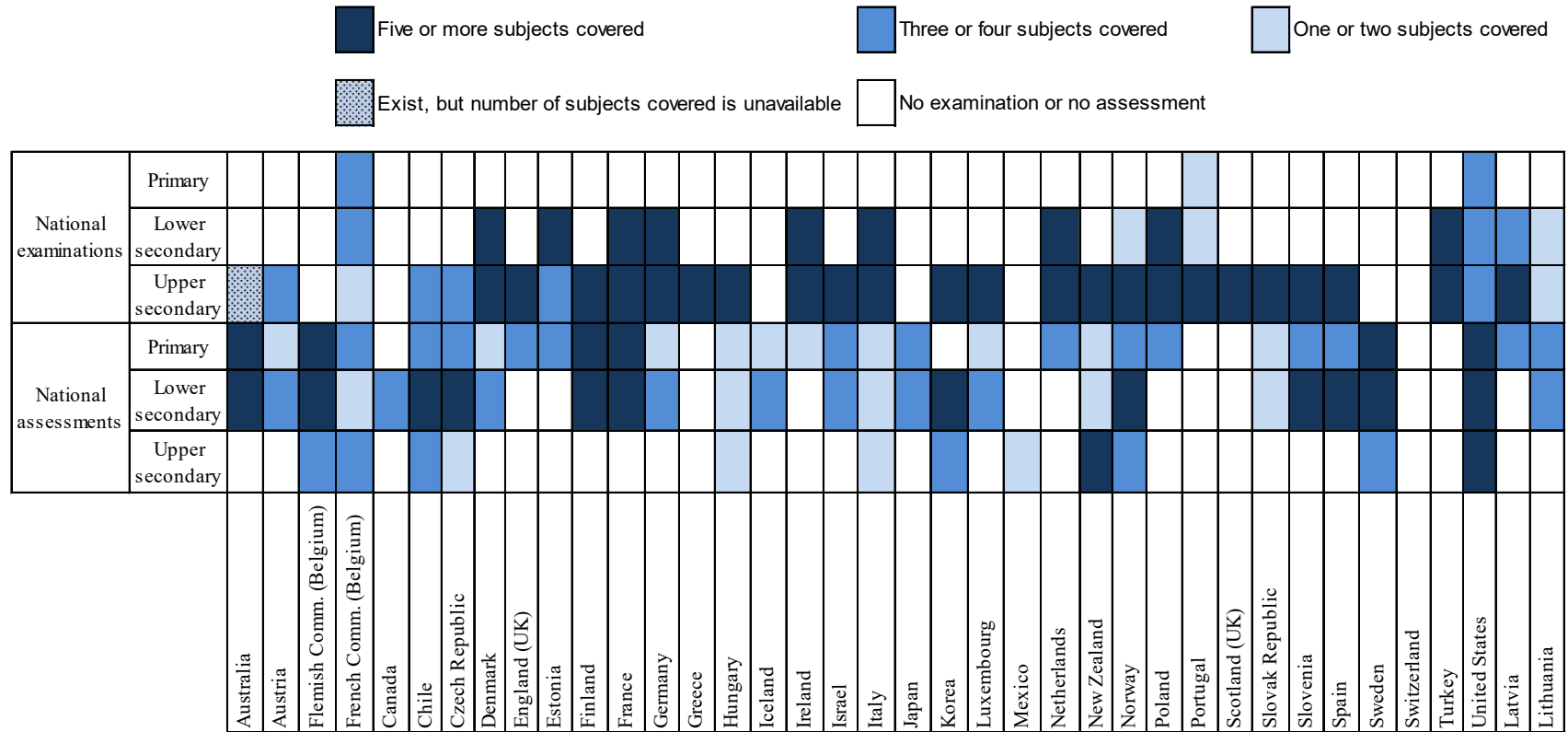
National assessments

Like national examinations, national assessments are centrally developed tests that are administered to students under fixed conditions. Unlike examinations, however, the primary purpose of a national assessment is to monitor outcomes to help evaluate the system. Kazakhstan's national assessment is the External Assessment of Academic Achievement (EAAA) and it is currently administered in Grades 4, 9 and 11 on a sample basis. The policy perspective on the national assessment discusses the EAAA in depth, but it will be mentioned in this policy perspective where it shares resources with examinations (e.g., the same item developers) or where their functions have overlaps and conflicts.

Key features of national examinations

National examinations are a pivotal component in several parts of a country's educational landscape. Results on examinations are often a key criteria in determining whether students are admitted to higher education and whether they receive scholarship assistance. Since examinations carry important stakes, they also influencing what students learn, what teachers teach, and how teachers assess if what they teach has been learned. The vast majority of OECD economies have national examinations at important decision-points for students, including 31 out of 38 at the end of upper secondary education (see Figure 3).

Figure 3. National examinations and assessments in public schools in OECD countries



1. Number of subjects covered in the assessment framework (subjects may be tested on a rotation basis). In many countries and economies, students take fewer subjects than are available in the assessment framework.

2. Data for the national examinations and assessments in Lithuania are drawn from authors' considerations based on OECD (2017^[12]), *Education in Lithuania*, Reviews of National Policies for Education, OECD Publishing, Paris. <http://dx.doi.org/10.1787/9789264281486-en>.

Note: Data were collected in 2015 and may not reflect changes since then.

Source: OECD (2015^[13]), *Education at a Glance 2015: OECD Indicators*, <http://dx.doi.org/10.1787/eag-2015-en>.

The extent to which national examinations fulfil their purposes depends upon the strength of certain test characteristics, which in turn are affected by several decisions related to the design and procedures of the tests. These elements are illustrated in Table 1 and discussed in the remainder of this section.

Table 1. Purposes, important characteristics and components of national examinations

	Purpose	Important test characteristics	Design and procedural components
Primary purpose(s)	Certify and select students	Integrity Reliability Validity (construct and content)	Test subjects Item types Testing mode Testing conditions
Secondary purpose(s)	Exert positive backwash effects		Marking Management and leadership

Purposes

National examinations serve two broad purposes. Their primary purpose is to certify that students have completed a level of education (e.g., upper secondary education), and/or to select students for future opportunities (e.g., scholarships or university admission). The secondary purpose of examinations is to exert positive backwash effects onto the education system.

In many countries, examinations are used in ways that go beyond their explicit purposes. A purpose refers to a deliberate function of the national examination. A use refers to how the results of an examination are employed. For example, some countries use examination results as an accountability measure for teachers, even though this is not a deliberate function of national examinations. Kazakhstan also uses examinations to serve several those outside of those mentioned in Figure 3, and this policy perspective addresses some of those uses as well.

Certify and select students

Certification refers to formal recognition that a student has met the basic minimum learning standards for a certain period of study. Selection refers to sorting students to determine which students are able to access a resource (Kellaghan and Greaney, 2019^[14]). One or several examinations are often administered as part of both certification and selection procedures. Depending upon the country, students' results on the examinations can contribute minimally or a great deal to the criteria that determine whether students are certified and/or selected.

Exert positive backwash effects

Given the important stakes of examinations, teachers and students feel pressured to orient their activities towards succeeding on them, a consequence known as a “backwash effect” of the examination (OECD, 2013^[15]). When directed properly, a backwash effect can be positive. Provided that the examination is aligned with the curriculum, examination pressure can encourage teachers to cover the curriculum and students to master it, a process that is known as reinforcing the curriculum (Bishop, 1999^[6]). Furthermore, the types of items on an examination can influence how teachers assess their students. For instance, an examination that contains multiple-choice, fact-based questions would incentivise teachers to use multiple-choice quizzes to determine whether students have memorised those same facts. On the other hand, an examination that assesses critical thinking through well-designed and non-predictable essay questions might motivate teachers to adopt similar assessment methods in their own practice. For these reasons, especially in countries with high-stakes examinations, attempts to change classroom assessment usually require changes in how students are assessed nationally (OECD, 2013^[15]).

Important characteristics of examinations

Integrity

Integrity refers to the degree that examination procedures are free from misconduct, such as cheating, dishonesty or tampering. A national examination is relied upon as a method of certifying and selecting students because it is expected to hold all students to the same standards and be administered and marked in a uniform manner. If there is doubt about the integrity of an examination, then these expectations are not met and the examination cannot be trusted to fairly certify and select students.

Reliability

Reliability refers to the consistency of a student's examination result; students who, in theory, repeatedly take a highly reliable examination should expect to receive a very similar result each time (American Educational Research Association, 2014^[16]). From a technical perspective, a reliable examination result is one with a low amount of statistical error. It is important for examinations to be highly reliable because decisions about a student can only be made confidently and efficiently if their examination result truly represents what they know and can do, as opposed to representing random chance or statistical noise. Broadly speaking, the importance of reliability (and integrity) increases with the stakes of the test, as even small differences in student results can have significant consequences.

Validity

A test's validity refers to what extent it assesses what it intends. Validity can be disaggregated into construct validity and content validity. Construct validity refers to the extent to which the test actually measures the skills or aptitudes (constructs) it sets out to. Content validity refers to the extent to which the test covers a representative amount of material associated with the construct (ETS, 2019^[17]). Validity can perhaps be best understood by reflecting upon what would be considered invalid. A test that intends to measure a student's ability in geometry has low construct validity if it asks questions about art. The same test is considered to have low content validity if it asks questions about squares but not circles. In the context of national examinations, a high level of validity means that the examination is assessing what the curriculum and national learning standards expect students to know and be able to do.

Risks

When these important characteristics are strongly present in an examination, then the examination is more likely to achieve its purposes. However, when examinations lack these important characteristics, not only will they struggle to achieve their purposes, but they might even create unintended, negative consequences. This section discusses what some of these risks can be.

Inefficient allocation of important resources

A lack of integrity and reliability in an examination would erode trust in the test and ultimately contribute to a misallocation of important resources. If an examination lacks integrity, a student's result could be a reflection of having received an unfair advantage. If an examination lacks reliability, a student's result could be a reflection of an individual marker's biases. In both cases, the student's result would not represent what he/she actually knows and can do. Consequently, underqualified students might receive valuable and limited resources, while the most deserving might not. In Georgia, widespread corruption around its national entrance examination awarded university places to students who paid bribes, which eroded public trust in the system and the value of a tertiary degree (Rostiashvili, 2004^[18]). Examinations with poor validity can also misallocate resources. These examinations assess students in areas that might have very little

relation to what they need to succeed in the future. When this occurs, resources can be given to students who are not prepared to take full advantage of them.

Distortion of teaching and learning

While examinations with high validity can help reinforce a country's curriculum and desired assessment practices, an examination with low validity can hold back these same efforts. If an examination has low validity, especially low content validity, it means that the examination is not assessing what is in the curriculum. Students would have difficulty preparing for the examinations through their normal classroom activities, and teachers cannot help students by merely following the curriculum. In these cases, teachers might deviate from the curriculum in order to teach students what they need to know to succeed on the examinations (Madaus, 1988^[19]). Similarly, teachers might have to develop new assessment methods, potentially at odds with those they are encouraged to use by the government, specifically to assess if students have learned material for the examination. When these activities occur, the examination is said to "distort" teaching and learning. In China, for example, the high-stakes national entrance examinations creates immense pressure for teachers to teach exam material, sometimes at the expense of developing other student aptitudes (Kirkpatrick, 2011^[20]). Such distortion can contribute to several, systemic consequences. The national curriculum might become difficult to implement because teachers feel pressured not to follow it. Similarly, new and encouraged practices might be difficult to embed because teachers will not have time to use them.

High-stakes examinations with low validity can also contribute to the emergence of a shadow education sector. Since students cannot prepare for the examination by studying the school curriculum, they might seek out additional opportunities outside of school, in particular private tutoring. This situation can contribute to educational inequity because it gives an educational advantage to students with more means (Bray and Kwo, 2013^[21]; Bray and Kobakhidze, 2014^[22]). Research from China, for example, suggests that private tutoring is more likely to benefit students from schools that already have more resources (Zhang, 2013^[23]).

Design and procedural considerations

Countries make several key decisions regarding how their national examinations are developed and administered. Ideally, all such decisions would be made to strengthen the integrity, reliability and validity of a country's examinations. However, countries must also consider their individual needs and resource limitations when making these decisions. This section discusses some of these trade-offs.

Testing conditions

Testing conditions include the resources that students are allowed to consult during testing, how closely monitored students are while testing, and even their physical comfort as they take the test. A common method of providing consistent testing conditions is to require that students take the test at designated sites. These sites are strictly monitored and their conditions are regulated. Operating enough testing centres around the country, however, is resource intensive. Where not all test-takers can be accommodated at testing centres, the examination will usually be administered at students' schools

Testing conditions can affect the integrity of an examination and the reliability of students' results. A regulated testing centre is more likely to consistently provide the resources that students need to test while preventing misconduct. Conditions at schools would vary more greatly, ranging from how closely students are monitored, and by whom, to how comfortable the test site is. Inconsistent testing conditions (e.g., too much noise or insufficient heating) can distract students, making their results partly a reflection of the circumstances under which they took the test (i.e., lowering reliability) and not just their ability.

Testing mode

Testing mode refers to in what format a test is administered. Historically, most tests have been administered on paper, meaning students receive test booklets and submit their responses on physical answer sheets. Advancements in technology have allowed for computer-based testing, where students receive test materials and submit their answers digitally². Where such technology is available in some parts of the country but not others, countries can also elect for a mixed-mode examination, where some students take the test via paper and some on computer.

The mode of a test can strongly affect its integrity. If examinations are administered on paper, there is greater chance that test materials can be “leaked” before testing occurs, or be printed with errors. Paper testing also requires human administrators, who can affect integrity by not strictly following testing procedures. Computer-based testing is generally considered to have higher integrity. Testing materials are kept digitally, without the need to print thousands of copies, which reduces the chance of accidental leaking. Computer testing also ensures that all testing follows the same rules.

Test subjects

The subjects that are assessed by an examination depend on the purpose of the examination. Since certification examinations assess their mastery of the national curriculum, the subjects that they assess also closely follow the curriculum. Selection examinations (assuming that a country has two examinations that serve different purposes) might assess different subjects because they aim to assess the competencies needed to succeed in the future, such as in higher education. However, it is likely that the national curriculum and national learning standards are designed to reflect to what students need to know and be able to do to succeed in the future. Therefore, the subjects assessed on a selection examination would be expected to closely reflect the national curriculum, though sometimes in a more general sense (e.g., reading literacy as opposed to language and literature).

What subjects are tested on national examinations strongly influences the validity of the examinations. A test that assesses all the material from a curriculum would have very high validity, but would be far too burdensome on students. Instead, countries evaluate the trade-off between validity, in particular content validity, and feasibility when determining how many and which subjects to assess.

Another method of improving validity is to differentiate between core and elective subjects, which allows students to define what is most valid with respect to their needs. Internationally, out of the 31 OECD economies with a national examination (see Figure 3), 23 offer five or more subjects, but this data reflects how many are offered, not how many one student takes. Internationally, countries have generally moved towards greater student choice in test subjects, which is reflective of more variety in upper secondary curricula that encourages students to pursue a diverse array of interests

A final method of improving validity is to offer different versions of the same subject so students with different interests can either demonstrate that they have mastered basic minimum competences or have acquired advanced understanding in an area. Mathematics is a subject that commonly features this type of arrangement. Students who wish to pursue further studies in mathematics might take an advanced version of the mathematics subject test, while others take a more basic version.

² The mode that test material is presented can also differ from the mode of students’ responses. For example, students’ could be given physical test booklets but submit their answers via computer.

Item types

Generally speaking, there are two different types of items that appear on national examinations closed-ended or fixed-response items (e.g., multiple choice), and open-ended or free response items. Open-ended items can range from filling in a blank, writing a formula or writing an essay.

The types of items that appear on an examination influence the test's construct validity. Even if a student is assessed in the same subject that appears in the curriculum, the type of item that a student responds to might not reflect the aims of the curriculum. A common example of this type of low construct validity comes from countries, such as North Macedonia, that have recently transitioned from a fact-based curriculum to a competence-based curriculum (Kitchen et al., 2019^[24]). Despite this transition, some items on the countries' national examinations might still assess students' ability to memorise information instead of asking them to demonstrate competences, which reduces validity. Open-ended items have the potential to more effectively assess some students' competences and higher-order skills (Tanujaya, Mumu and Margono, 2017^[25]), but these items are more difficult to develop than closed-ended items, and there are trade-offs associated with the marking of such items, which are discussed next.

Marking

Related to item types is how student responses are marked. Closed-ended items are marked more reliably than open-ended ones, which require a degree of subjective interpretation. Like the mode of testing, marking can also be performed by humans or electronically. Tests that are administered on paper can still be marked digitally, such as by scanning student answer sheets through a machine, a process referred to as optical mark recognition. Likewise, tests administered on computer can still be marked manually, in particular many open-ended items.

How student responses are marked is strongly related to the reliability of student results. Human marking is considered less reliable than electronic marking, even for closed-ended items (Leiva, Ríos and Martínez, 2006^[26]). Marking also affects the integrity of a test. If human marking is required, students' answers (especially if the test is paper-based) must be transported and seen by potentially many markers, which increases the risk of misconduct.

For these reasons, introducing open-ended items to national examinations, while potentially very useful, is also often controversial, especially in societies where there is low trust and/or the examinations are perceived to have very high stakes. In Japan, long-standing plans to introduce open-ended items to the national university entrance examination were further delayed because a pilot test revealed several marking inconsistencies (Japan Times, 2019^[27]).

Management and leadership

How examinations are managed is reflective of the national context and produces different sets of trade-offs. At a macro level, countries can manage examinations via three models — market, quasi-market and nationalised (Opposs et al., 2020^[28]). The type of model that is chosen affects the level of government accountability of the responsible agency(ies), and the extent to which the agency is able to exert its own agenda over the examinations. In a national model, for example, acting in the interest of the government is assured because the bodies involved are under the direct control of the government. However, the direct control limits the agency's ability to exercise its technical expertise to help guide policy.

Regardless of the type of model, internationally almost all bodies that manage examinations are led by a management board, such as the Singapore Examinations and Assessment Board and the Hong Kong Examinations and Assessment Authority. Board membership is designed to be representative of all stakeholders and is typically composed of the chief executive of the examinations agency, ministry of education representatives, higher education representatives, and representatives from the private and non-profit sectors. Although boards are responsible for setting the strategic direction of examinations

agencies, their members are not necessarily examinations experts. To ensure that their decisions are technically sound, they are typically advised by a number of technical and expert committees who provide guidance related to education measurement and subject domains. For instance, the New Zealand Qualifications Authority has seven such committees (New Zealand Qualifications Authority, n.d.^[29]) and the Hong Kong Examinations and Assessment Authority has one for each tested subject (Tong, Lee and Luo, 2020^[30]).

The examinations agency itself acts as a secretariat. It is directed by its management board and employs highly qualified technicians to execute the board's vision. Nevertheless, given the familiarity of the examinations agency with their products, it is important that it is able to influence its own strategic direction and weigh in on important decisions related to its tests. This type of input is usually provided through the agency's representation on the management board.

National examinations in Kazakhstan

This policy perspective has thus far discussed key features of the education system of Kazakhstan that are related to national examinations, and introduced a framework for speaking about national examinations in general. This section uses that framework to analyse Kazakhstan's national examinations.

Kazakhstan has two national examinations, one for upper secondary certification and another for tertiary entry

In 2004, the Unified National Test (UNT) was created with the dual purpose of certifying students' completion of upper secondary education and selecting them into university and for related scholarships. An important factor for the creation of the UNT was to improve the integrity of the university admissions process. Before the UNT, universities were largely responsible for developing their own examinations and admissions criteria, both of which were opaque, vulnerable to misconduct, and contributed to inequity (OECD/The World Bank, 2007^[31]).

While reception to the UNT was very positive, over the years the design of the UNT became less fit-for-purpose vis-à-vis the modern educational needs of the country. Two concerns in particular have been highlighted. First, as noted by several OECD reviews, items on the UNT encouraged students to memorise facts instead of demonstrate applied learning and higher-order thinking (OECD, 2014^[9]; OECD, 2017^[11]; OECD, 2007^[32]). Second, the stakes of the UNT, amplified by the limited availability of higher education spaces and funding, were creating too much pressure for teachers and students (Jumabayeva, 2016^[33]).

In response to these concerns, and consistent with OECD recommendations, the UNT has undergone several reformations. The number of mandatory subjects was reduced and the number of times that students are allowed to take the test was increased. In 2017, Kazakhstan introduced the Grade 11 Final Attestation, which certifies students as having completed upper secondary education. The UNT is now only used for selection purposes.

Unified National Test

Primary purposes

The UNT selects students from Kazakhstan to enter university and to receive scholarships. While completion of upper secondary education is a requirement in both cases, there is no weight given to students' marks, except indirectly through consideration of Altyn belgi awardees (see the description below in the section about the Final Attestation). Thus, the UNT is the only standardised, competitive criteria in both cases, though threshold scores on the UNT can vary depending upon the needs of certain fields (OECD, 2017^[11]). The maximum score on the UNT is 140 points, while at least 50 points is required to

enter university and be eligible to receive a scholarship. According to data provided by Kazakhstan, roughly 78% of Grade 11 general upper secondary students took the UNT in 2019 and 75% of those who did passed.

Design and procedural considerations and their effects on important test characteristics

Important design and procedural considerations of the UNT are illustrated in Table 2.

Table 2. Design and procedural considerations for the Unified National Test

Topic	Specifications	Notes
Testing mode	Paper-based (soon to be computer-based)	Can be taken multiple times in grade 11 (soon to be grade 12)
Testing conditions	Administered at the regional centres of the National Testing Centre, not in students' schools	All administration is overseen by several bodies, including the regional Quality Control Committee
Test subjects	Language of instruction, mathematics and the history of Kazakhstan are compulsory. Two electives are chosen by the student.	While two electives are chosen, students can only select them in combination, not individually. In general, according to the National Testing Centre (NTC) website, the combinations are divided into sciences (e.g., mathematics and physics) and humanities (e.g., a foreign language and world history). Possible combinations are determined by the type of upper secondary school a student attended
Item types	All items are close-ended, multiple choice (single answer or multiple answers)	Kazakhstan is planning to introduce open-ended items to the UNT in order to improve its capacity to assess more complex skills
Marking	Occurs electronically	Results are made available online and can be retrieved the day of the test
Management and leadership	A working group formed by the Division of Higher Education within the Ministry of Education leads the UNT. The working group is composed primarily of university representatives. The National Testing Centre administers the UNT, but does not lead its development	-
Use of results (aside from deliberate purposes)	System monitoring School audit criteria (maybe teacher appraisal criteria)	UNT results are included in annual reports produced by the Information Analytics Centre. Students can access their results online and schools also receive the results of all their students

Integrity

Ensuring the integrity of the UNT is a high priority. Centre-based administration strengthens UNT integrity considerably. The speed with which results are returned is important to stakeholders because it suggests that there was no human interference in determining a student's results.

Reliability

The reliability of the UNT is very high, also owing to the fact that all questions are multiple-choice and marked digitally. However, there are plans to introduce open-ended items into the UNT in order to better assess higher-order skills. Introducing these items could improve the validity of the UNT, but would necessitate human marking, which could make students' results less reliable than they are now.

Validity

The poor validity³ of the UNT has been a long-standing concern. Regarding construct validity, the UNT focuses almost exclusively on assessing the amount of knowledge a student has acquired, which might

³ In this policy perspective, the validity of the UNT is discussed in terms of its construct and content validity vis-à-vis national learning expectations, not its fidelity with respect to its test development specifications, which is very strong.

not be a useful measure to universities who wish to select the students most capable in the areas of critical thinking and problem solving. The OECD examined a sample of items from the UNT as part of this review and confirmed that the items still focus strongly on assessing students' memorisation of specific facts. For example, mathematics questions assess whether students can recall complex procedures to solve logarithms and matrices and history questions ask students to read a law and identify from which civilisation the law originated. In comparison, there is a lack of tasks that ask students to apply their knowledge to solve problems in real life contexts. Concerns about validity have only become magnified since the introduction of the new competence-based curriculum, which emphasises critical thinking and problem solving. There are plans to introduce new items types into the UNT in 2021, which aligns with when the first cohort of students who studied the new curriculum will graduate from upper secondary school.

Regarding content validity, the range of subjects assessed by the UNT is relatively narrow compared to international benchmarks. Students can only take one combination of electives and, while there are different versions of the test for some subjects, students cannot choose which version to take. As a result, students have difficulty demonstrating their full range of the skills and aptitudes through the UNT.

Evidence suggests that the UNT's lack of validity might be contributing to students seeking out shadow education opportunities. PISA 2018 data shows that 40% of 15-year-old students from Kazakhstan attend enrichment courses in reading, compared to 15% across OECD countries. Of these students, 62% from Kazakhstan attend enrichment courses do so to prepare for examinations, compared to 45% across OECD countries (OECD, 2019^[31]).

Furthermore, the UNT's lack of validity limits its capacity to reinforce the use of the new curriculum and criterion-based assessment. The UNT encourages teachers to help students memorise facts instead of develop their higher-order skills, and motivates teachers to focus on the relatively narrow amount of content that is assessed by the UNT. These circumstances distract teachers from teaching the full breadth of the curriculum and assessing students based on what they can do vis-à-vis the national standards.

Final Attestation

Primary purposes

The Final Attestation, along with student marks, certifies completion of general upper secondary education. Like a student's marks in general upper secondary school, the results of the Final Attestation are given on a 5-point scale for all subjects. To graduate, students' must achieve a three or higher on all their school marks and all their Final Attestation subjects. Final Attestation data are kept locally, and thus national results are unavailable, but it is widely acknowledged that almost all students pass the Final Attestation.

In addition to certifying completion of upper secondary education, the Final Attestation contributes to the selection of students for the "Altyn belgi" award. Students who receive a mark of five in all their school marks from Grade 5 through Grade 11, and who achieve a mark of five in all Final Attestation subjects, receive this award. In addition to national recognition, Altyn belgi designation acts as a tie-breaker in state scholarship eligibility. If two students achieve the same score on the UNT, the Altyn belgi student receives priority consideration. In 2018, roughly 3% of Grade 11 graduates received this designation.

As mentioned previously, students in colleges do not take the Final Attestation. Therefore, almost 40% of the upper secondary school population do not receive a nationally standardised certification, nor are they eligible to be receive Altyn belgi awards.

Design and procedural considerations and their effects on important test characteristics

Important design and procedural considerations of the Final Attestation are illustrated in Table 3.

Table 3. Design and procedural considerations for the Final Attestation

Topic	Specifications	Notes
Testing mode	Primarily paper-based, along with other formats (see Table 4)	-
Testing conditions	In each school, the principal creates a Final Attestation commission, composed of teachers, deputy principals, public organisation representatives and representatives from the parents' committee. The commission administers the test.	Test materials are developed centrally and sent to schools along with guidance and manuals regarding administration. There is no formal oversight of test administrations.
Test subjects	See Table 4	-
Item types	Closed-ended, open-ended long-form tasks, essays, oral answers (see Table 4)	-
Marking	By a school-level commission	Results are marked by the commission on the day of testing and announced the following day Results for students who are eligible for <i>alтын белгі</i> consideration must be verified by the <i>rayon-level akimat</i> (local-level government body).
Management and leadership	The Committee for Early Childhood and Secondary Education within the Ministry of Education leads the Final Attestation. National Testing Centre (NTC) provides technical advice and support and develops materials.	-
Use of results (aside from deliberate purposes)	-	Results are sent to the <i>rayon-level akimat</i> , but are not transferred to a central authority.

Table 4. Subjects and administration formats of the Final Attestation

Subject	Format
Language of instruction and literature	Essay
Algebra and analysis	Long-form answers to tasks, such as writing formulas or drawing charts. Students from schools that specialise in mathematics perform six tasks, while other students perform five
History of Kazakhstan	Oral presentation and question/answer
Kazakh or Russian as a foreign language	Multiple-choice closed-ended
A combination of two electives (Physics, Computer Science, Geometry, Chemistry, Biology, Geography, World History, Literature, Foreign Language)	Multiple-choice closed-ended (specific questions for some subjects can vary depending upon if a student attends a school that specialises in mathematics)

Integrity

The integrity of the Final Attestation is weaker than that of the UNT. All general upper secondary Grade 11 students must take the Final Attestation, and therefore the resources needed to administer the test are greater. Capacity at National Testing Centre (NTC) regional centres is insufficient to support such a large demand, so the Final Attestation is administered at the school-level. Likewise, there is no oversight by the Quality Control Committee, who do not have the capacity to monitor all administrations of the test.

To bolster the integrity of the Final Attestation, Kazakhstan develops the test centrally and issues the test with strict protocols and guidance. For example, all testing is to occur at the same time on the same day and materials opened according to a strict schedule. Marking the results the same day of testing and announcing the results the following day also helps communicate that there was little opportunity for misconduct to occur. Nevertheless, these efforts do not fully compensate for the absence of standardised testing conditions.

Reliability

The reliability of the Final Attestation is relatively low. School-level administration implies that tests cannot be administered on computer because many schools do not have the infrastructure to support it. The test's items, many of which are open-ended, also cannot be marked digitally. Furthermore, marking for all test items is performed by different people in each school and without moderation by external parties, except in the case of Altyn belgi candidates.

Validity

A significant reason for the creation of the Final Attestation was to introduce an examination with a high level of validity vis-à-vis the curriculum and learning standards. For this reason, the Final Attestation items include an essay, an oral history test and open-ended tasks on algebra and analysis. These items have the potential to strengthen the Final Attestation's construct validity in that they encourage students to apply what they have learned instead of just demonstrate that they have acquired a certain set of facts. Like the UNT, however, there are concerns related to the content validity of the Final Attestation. Students can only choose one elective combination and, while students can take different versions of some subjects, which ones they take are prescribed for them instead of chosen by them. These circumstances limit the extent to which the Final Attestation can assess of students' mastery of the entire breadth of the curriculum.

Risks

Final Attestation results carries little meaning or value for students

The weak oversight and moderation around Final Attestation contributes to students receiving inflated marks that carry little meaning. A general upper secondary qualification, certified by the Final Attestation, does not carry much weight because the public does not trust the result of the test, which could contribute to the higher unemployment rate and lower earnings of individuals who hold only this qualification. The fact that Altyn belgi candidates' results receive external moderation contributes to inequity because all students deserve the right to have a reliable certification of their performance, not only the highest achieving ones.

As mentioned previously, national examinations are often used in several meaningful ways that are outside of their explicit purposes. In North Macedonia, the *state matura* (end of upper secondary examination), serves as a system monitoring tool. In the United Kingdom, results on the A-levels examinations can be submitted by job candidates when searching for employment. The Final Attestation cannot be used in these ways, or others, because students' results on the tests are too unreliable. As a result, Kazakhstan is missing out on opportunities to make its assessment system more efficient (see policy perspective on the national assessment) and generate more useful information with which to make important decisions.

The Final Attestation exerts few positive backwash effects on the education system

A prerequisite for exerting backwash effects, whether positive or negative, is that the test carry stakes. The results of the test need to matter for students and teachers so they feel motivated to apply themselves in ways that the test demands. Since passing the Final Attestation is nearly assured, it does not place pressure on students to pass it, or on teachers to help students prepare.

Other uses of Kazakhstan's national examinations

Kazakhstan's examinations are also used in ways outside of their explicitly designed purposes. Given concerns about their characteristics, however, many of these uses are inappropriate and might create negative consequences.

Unified National Test results are used as part of school audit procedures and teacher appraisal criteria

Kazakhstan has a strong system of school audits (see policy perspective on school evaluation). One of the components of this system is known as preventative control, which utilises a risk assessment to determine if schools should receive a visit by representatives from the Quality Control Committee. Since Kazakhstan lacks reliable school-level data about student learning, a school's un-contextualised UNT results are part of the risk assessment criteria that determine school visits.

Additionally, Kazakhstan's teacher appraisal system is fragmented and each school has the flexibility to set its own in-school criteria for teacher attestation (see policy perspective on initial teacher education). There are no data that aggregate all schools' criteria, but previous OECD reviews noted students' results on the UNT are teacher appraisal criteria in some schools. These uses of the UNT help distort examination and assessment practice by incentivising schools and teachers on the top students instead of the progress of all students

Unified National Test results are used to monitor the performance of the education system

Even though an unrepresentative student sample takes the Unified National Test (UNT), results from the UNT are one of the most commonly used indicators in Kazakhstan's system monitoring efforts. Annual reports produced by the Information Analytics Centre, a high-capacity data collection and analysis agency, disaggregate UNT results according to oblast, gender, language of instruction and other dimensions. NTC's website also releases large amounts of UNT data. While not in the form of reports, users can download spreadsheets of UNT data by year, subject and oblast. Until 2017, the 100 schools with the top average UNT scores were also identified on the NTC website, which further contributed to an inequitable focus on the highest performing students

Management and leadership

Kazakhstan has a national model of examinations management with NTC being a subsidiary of the Ministry of Education. As is the case with national management models, NTC is highly accountable to the government but has a limited mandate to contribute to policy-making and is considered an "executor of plans". NTC does not have its own management board or advisory committees and thus lacks a strong voice in the policy arena, where its contributions are usually limited to technical advice. Since NTC has a limited a public policy role, its tests are not steered or conceptualised by them. As indicated in Table 2 and Table 3, a working group in the Division of Higher Education leads the UNT, while a separate working group in the Committee of Early Childhood and Secondary Education leads the Final Attestation.

To execute the country's national examinations and assessments, NTC employs almost 500 staff in its central office and in testing centres around the country. To develop items for the UNT and the national assessment, the External Assessment of Academic Achievement (EAAA) (the ministry develops items for the Final Attestation according to specifications that NTC makes), NTC taps into a registry of over 2 500 item writers. These individuals are provided with training, test specifications and an item development manual. Each item that is written is reviewed by a reviewer and piloted before use in a test. NTC produces a report about item functioning following pilot testing, and following actual testing it produces the aforementioned reports about UNT results. Annually, roughly 350 item writers are contracted for UNT development, 130 for the EAAA and 200 for the Final Attestation.

Kazakhstan has recently begun an education modernisation project with the support of the World Bank. A large component of this project is providing technical assistance to NTC around strengthening examinations. The World Bank plans to identify and train around 500 item writers who will develop items for the UNT and EAAA.

Review of the context

National examinations serve vital roles in helping a country achieve its educational and economic goals. Specifically, they allocate valuable and limited resources and help reinforce the teaching and learning aims of a country. To achieve these functions, national examinations must be designed to have high levels of integrity, reliability and validity. The extent to which Kazakhstan's national examinations are designed to strengthen these characteristics and fulfil their purposes forms the basis of the analysis and recommendations found in this policy perspective.

A fundamental issue regarding Kazakhstan's examinations is that how they should be used is not clearly defined, and therefore they are being used in ways for which they are not well suited. For example, the UNT is used to monitor the system, even though it collects data from an unrepresentative (i.e., the highest performing) sample of students. Continued usage of the UNT in this way thus contributes to keeping the educational focus on the best students and not on helping all students meet basic minimum standards.

The primary purpose of Kazakhstan's national examinations is to certify and select students, which requires that the exams have high degrees of integrity, reliability and validity. Regarding the UNT, its current levels of integrity and reliability are fairly high, but proposed developments could alter this situation. There are, for instance, encouraging plans to introduce open-ended items to improve the test's construct validity. However, introducing such questions would require human marking and, without careful planning, could affect the reliability of the results. On the other hand, the integrity and reliability of the Final Attestation are relatively low, which prevents the results from being taken seriously and limits the value of the certification it confers.

The secondary purpose of Kazakhstan's national examinations is to exert positive backwash effects on teaching and learning. This purpose is not being fulfilled well by either examination, which is complicating the implementation of Kazakhstan's new curriculum and criterion-based assessment. The reasons that examinations are not achieving this function differ. With respect to the UNT, the primary limitation is poor validity. UNT items are not well aligned with the curriculum, which prevents the UNT from adding educational value and distorts the activities of teachers and students in negative ways. With respect to the Final Attestation, its low integrity and reliability leads to a general belief that the test does not matter much, which limits its influence on teacher and student behaviour.

A final issue concerns how the examinations are managed. NTC has technical expertise and familiarity with the examinations, but does not have a leadership role over the products it creates. As a result, management over NTC and its examinations, despite their overlapping purposes, is separated between the Committee and two different ministry divisions, each of which operates with minimal technical advice. Without common, representative and informed leadership, important aspects of the tests are not considered when decisions are made, which in turn prevents the tests from performing all their functions.

Recommendation 1. Clearly define the purposes and uses of national examinations and assessments

In Kazakhstan, there is a general awareness of the different purposes served by the national examinations, and how they differ from those of the national assessment (EAAA). Nevertheless, the UNT is used in ways for which it is not suited. While there are important, direct reforms to the examinations that should be considered, the effectiveness and sustainability of those reforms require that they be aligned and coherent with each other. This recommendation suggests that a national assessment framework be developed to act as a policy guide and serve as a conceptual reference for establishing a clear and shared understanding of examinations and assessments in Kazakhstan.

1.1. Develop a national assessment framework that defines the purposes of national examinations and assessments

Evidence

It is not unusual for a national examination to be used in diverse ways, but their usage needs to be appropriately supported by their design for them to be effective in those roles. In the case of the UNT, its design does not adequately support all the ways in which it is used. Since 2017, the UNT has been taken by a voluntary sample of students, making it unrepresentative at any level of analysis, and there have always been concerns over its validity, which make it a poor measure of educational attainment vis-à-vis the curriculum. However, UNT results are still prominent in system monitoring reports and are used to understand the performance of different regions in the country. Moreover, school and teacher accountability (i.e., preventative control and teacher attestation) is also directly informed by these results, which can significantly distort how schools and teachers behave (e.g., they could discourage low-performing students from taking the UNT in order to boost average UNT performance).

Using examinations in inappropriate ways can also have indirect effects. By trying to make the UNT an effective monitoring tool, the government might make decisions that negatively affect the examination's ability to select students, which is its chief function. For example, developing questions to capture student performance at the lower end of the student distribution, which is necessary for a monitoring tool, might prevent the test from accurately discriminating performance at the upper end of the student distribution, which is necessary for a selection tool. Also, the UNT's use in system monitoring detracts from and risks distorting the EAAA's role. For instance, stakeholders might conflate the purposes the UNT and the EAAA, which is likely contributing to a current false perception that the EAAA has stakes and affects the utility of that test (see the policy perspective on the national assessment).

To co-ordinate the development of different assessments and clearly communicate their different purposes, many countries develop a national assessment framework, which is a document that outlines what national assessments are to be administered and why. The assessment framework usually discusses national assessments and examinations in the context of overall assessment goals and is aligned with the national curriculum framework to promote national learning objectives (Asian Development Bank, 2017^[34]).

Kazakhstan does not have a national assessment framework. The purposes of each national test are defined in separate sets of national legislation, with a different law governing each test. Since understanding of the national tests is developed separately at different times, it is not surprising that that understanding of their roles is somewhat unclear.

Recommended actions

Create a national assessment framework

The OECD recommends that Kazakhstan develop a national assessment framework that is based on the principles found in the national curriculum, in particular those of competency-based education and criterion-based assessment. The resulting framework would describe how the assessment tools align with the curriculum to support student learning. It would also describe the purposes of all student assessment across the national system, and define the contribution of each national assessment and examination, clearly setting out the distinct purposes of each. Specific sections that should appear in Kazakhstan's national assessment framework include:

- Clear vision statements that are focused on student learning and linked to an understanding of the different purposes of assessment
- An overview of how assessment supports other aspirations for the school system, such as the competence-based curriculum and school review

- A explanation of each assessment (both centralised and classroom), its purpose, when it will be administered, in which subjects, and how its collected data are used and reported
- A description of the levels of proficiency that students can occupy at the end of each grade and what is required for students to meet expectations (much of this is already described elsewhere, in particular in materials related to criterion-based assessment, and can be brought into the national assessment framework to centralise the information)

It should be noted that a national assessment framework is not intended to replace the specifications of individual examinations and assessments, which are much more technically focused on an individual test. However, it is designed to act as a reference point for those documents so the resulting tests are coherent with each other and aligned with national learning objectives. Box 2 provides an example of how the national assessment framework is being developed in South Africa, which has recently decided to co-ordinate several different national assessments through a single framework.

Box 2. The National Integrated Assessment Framework of South Africa

The education system of South Africa has historically conceptualised its national examinations and assessments separately. Its national assessments (Annual National Assessments) were designed and managed separately from its national examinations. Importantly, both were conceived separately from regular classroom assessment and the links between the external assessments and how teachers assessed their students were limited.

Recognising the need to align all assessment efforts, in 2018 the South African Department for Basic Education began implementing the National Integrated Assessment Framework (NIAF). The NIAF outlines three complementary assessment programmes: systemic assessments, diagnostic assessments and summative assessments. There is an emphasis on using the three programmes to support teacher practice. For instance, reports based on national examinations and assessment results should be comprehensive and tailored to a target audience. The purposes, role and process of programme are:

- The systemic assessment (national assessment) evaluates the overall education system. NIAF stipulates that this test will be sample based, when it will be administered and to which grades. The systemic assessment is also linked to participation in international assessments and surveys (TIMSS, PIRLS, SACMEQ, TALIS) to ensure complementarity between the instruments and avoid overlap.
- The diagnostic assessment programme aims to support and strengthen teachers' classroom assessments and to enable them to identify learning gaps. It consists of giving teachers assessment tools, manuals, digital applications, and exemplar tests and test items.
- The summative assessment programme (national examinations) consists of end-of-year examinations in selected subjects and grades and helping teachers in accessing high-quality test items.

Sources: Government Communication and Information Systems Republic of South Africa (2018^[35]), Official guide to South African 2017/2018 Education. <https://www.gcis.gov.za/sites/default/files/docs/resourcecentre/pocketguide/08-Education-1718.pdf>; Department Basic Education Republic of South Africa (2018^[36]) Annual performance plan 2018/2019.

Stipulate that national examinations should not be used as system monitoring or school and teacher accountability tools

An important consideration when developing the national assessment framework is to reconcile the system monitoring responsibilities of Kazakhstan's national examinations and assessments. The EAAA is Kazakhstan's principle system monitoring tool, and this role should be made explicit in the national

assessment framework. The UNT is not representative of the Grade 11 population and is therefore ill suited to monitor the system. Additionally, a key test characteristic of system monitoring tools is stable, reliable measurement over time. Many national examinations, including the UNT, are not designed for results to be comparable across different administrations, further making them inappropriate system monitoring tools. However, because the UNT is used to allocate scholarships and university places, its results can be used to measure how equitably these opportunities are being allocated on a year-to-year basis. For instance, are students from rural and urban areas receiving scholarship assistance at similar rates, or students in upper secondary schools compared to students in technical and vocational colleges?

While the Final Attestation is administered to almost all general upper secondary students, its low integrity and reliability prevent it from being a system monitoring tool. If these item characteristics are strengthened, as suggested in Recommendation 2.2. , then Kazakhstan can consider using the Final Attestation in this way (though the primary reason for strengthening these characteristics is to improve the educational value of the Final Attestation, which is discussed in Recommendation 3.3.).

Recommendation 2. Maintain and improve the integrity, reliability and validity of the national examinations

The UNT was designed specifically with the intent to be trusted and reliable, though its validity is a concern. Proposed changes that are designed to strengthen its validity might weaken its integrity and reliability, and this trade-off needs to be carefully managed. Regarding the Final Attestation, its integrity and reliability are relatively weak, and the lack of these characteristics prevents the Final Attestation from being taken seriously and used for important functions that can help improve student learning.

2.1. Introduce open-ended items gradually into the UNT

Evidence

Internationally, many countries are introducing open-ended or constructed-response items into their national tests (OECD, 2013^[15]). While open-ended items have the potential to assess different types of higher-order skills, they are also more difficult to design, mark and code reliably than closed-ended, multiple-choice items. Furthermore, in countries with high-stakes testing cultures, closed-ended items are considered more trustworthy because they are marked by computers and results are returned quickly, which alleviates public concerns about corruption. For these reasons, it is important to consider how high degrees of integrity, whether real or perceived, and reliability can be maintained when introducing open-ended items for the first time into a high-stakes examination.

Countries have taken different approaches to deal with this issue. In Turkey, open-ended questions and other question formats were trialled in the new National Assessment of Student Learning (ABİDE) before they were incorporated into the country's high-stakes central examination (Kitchen et al., 2019^[7]). In the United States, the Race to the Top programme, which is aimed at measuring 'real student knowledge and skills', encouraged the development of new assessments that included open-ended items in English and mathematics. However, that stage was reached after three years of test design and development and an 18-month field trial involving over a million students (Pearson, 2018^[37]).

Kazakhstan plans to introduce open-ended items into the UNT. These items could help measure the diverse, higher-order skills that are present in the curriculum. Nevertheless, Kazakhstan has a very high-stakes examinations environment and a history of public suspicion of examination results. Introducing open-ended items will undoubtedly create controversy and Kazakhstan should only do so if it can ensure that the items can be marked securely and reliably. Currently, however, there are no plans for how to introduce the items in a methodical manner that maintains public trust in the examination.

*Recommended actions***First introduce simpler open-ended items and gradually introduce more complicated ones after thorough testing**

While open-ended items are typically thought of as asking students to input freely constructed text, they actually occupy a large range of item types. This distinction is important to make because different types of open-ended items have different levels of reliability. For example, a long-form essay can be difficult to rate consistently, but whether or not students spelled a word correctly can be easier. Technology is also helping to improve the reliability of open-ended items. One type of open-ended item that has been recently developed are technology-enhanced items. These questions ask students to interact with different components on the screen, the process of which generates data that constitutes the student's response. Examples of these items are asking students to drag and drop, categorise or put objects/text in a certain order. These types of items, though open-ended in nature, can still be marked electronically, which strengthens the reliability of students' results (Pearson, 2018^[37]).

Kazakhstan plans to transition to computer-based administration for the UNT. When this occurs, the country should first introduce technology-enhanced items, as these have the potential to assess a greater variety of student skills while limiting the trade-off in reliability. If computer-based testing is still not planned for several years, and introducing open-ended items is a priority in the meantime, Kazakhstan should consider first introducing simpler open-ended items and, once those are tested, established and trusted, gradually test more complicated ones for later use.

Begin a campaign to build understanding about the new items and maintain trust in the examination

Introducing open-ended items can strengthen the validity of a test, but can also affect public trust in the examination. If open-ended items are introduced without building sufficient understanding of their importance, and trust in their reliability, then students and universities might become disenchanted with the system and seek out alternative methods of signalling aptitude (i.e., a separate entrance examination for each university or attending university abroad). This situation could erode the rigour (both real and perceived) of the education system, and lead to greater inequity as students with the most means can more easily access and prepare for alternative assessment schemes.

The United Kingdom commissioned a two year study about how to disseminate information in order to improve understanding of and trust in assessments (Chamberlain, 2013^[38]; Simpson and Baird, 2013^[39]). Three general recommendations were formed:

- Develop educational assessment frames (a means to simplify and contextualise technical information through metaphors, examples, or careful sound bites)
- Use applied and interpretive information, not description and explanation, to help users understand not just what the information is, but how it affects them
- Recruit influential peers from assessment stakeholder groups as information brokers so users hear from people at the “grass roots level” to whom they can relate

Kazakhstan should build an information campaign around the introduction of open-ended items (especially technology-enhanced items, which can create confusion) and high-quality assessment in general. NTC and the Information Analytics Centre are well positioned to provide information about assessment. Akimats are well connected to the schools in their rayons and oblasts and can act as brokers of the information.

2.2. Introduce oversight into Final Attestation procedures to improve the value of the examination and prepare it to serve more functions

Evidence

A certification examination serves to signal that a student has mastered the basic minimum competencies that they are expected to have acquired in upper secondary education. For this certification to be valuable, however, the integrity and reliability of the examination must be high so employers trust student results' to accurately signal what they know and can do. As mentioned previously, examinations that are administered at the school-level must take extra precautions to ensure high degrees of these important characteristics.

In Queensland, Australia, school-based assessments are developed, administered and marked at the school-level. To maintain consistency across schools, all tests and their marking schemes vis-à-vis national standards must be approved by the Queensland Curriculum and Assessment Association (QCAA). After students have taken their tests, a sample is sent to QCAA for validation (Baird et al., 2018^[40]). In Hong Kong, district co-ordinators are responsible for supporting teachers who administer and mark school-level examinations (Tong, Lee and Luo, 2020^[30]). In Romania, many classrooms have camera surveillance to help maintain procedural integrity (Kitchen et al., 2017^[41]).

In Kazakhstan, although the Final Attestation nominally certifies completion of upper secondary education, the signal that it sends has little value in society and the economy. Students who do not wish to continue with their education struggle to find good work with their general upper secondary qualification (this situation is also indicative of a weak workforce training sector). Part of this challenge is related to the integrity of Final Attestation procedures and the reliability of its results. There is no external moderation of test administration, so there is no guarantee that students are taking the test under similar conditions. Much of the test consists of open-ended, free response items, but there is no external verification of marking (e.g., validation from an agency or peer marking) unless students are Altyn belgi candidates. Consequently, employers, and society in general, have difficulty trusting that certification conferred by the Final Attestation truly reflects what students know and can do.

The lack of integrity and reliability is not only affecting the value of the Final Attestation's certification, but is also preventing it from serving additional, useful functions. Internationally, some national certification examinations serve system monitoring purposes because all students at the end of upper secondary education take it, which eliminates the need for a national assessment to be administered that same year. As Figure 3 shows, far fewer OECD economies have a national assessment at the end of upper secondary education than a national examination (12 compared to 31). Moreover, out of the 12 that do, eight either do not have a national examination or test fewer subjects on their national assessments compared to their national examinations. Even though the Final Attestation is relatively well aligned with the curriculum, it cannot be used for monitoring purposes because its lack of procedural integrity and low reliability. These circumstances necessitate that the EAAA be administered in Grade 11, meaning that most students take three standardised tests during their final year of schooling.

Recommended actions

Broadly speaking, Kazakhstan should introduce significant moderation into the development, administration and marking of the Final Attestation. These measures could include, but are not limited to:

- Random external visits during administration – These visits could be conducted by the Quality Control Committee or the local akimat to validate test administration. Schools to be visited could change from year to year, but would not be announced prior to testing.
- Peer moderation of marking – Within schools, testing committees can establish regulations for checking the marks that students receive. For example, a teacher who is unassociated with the committee and the administration of the Final Attestation can be asked to mark student responses.

This teacher's mark can be compared to the mark a student received from committee members to determine if there are discrepancies that need to be addressed. In some cases, peer moderation could also occur across schools (e.g., testing committees from different schools provide marking support to each other).

- Spot checks of a sample of student work – Regional NTC staff or local akimat staff could receive a random sample of student work every year to mark according to marking guides. Schools whose marks are found to deviate significantly from external marking can receive additional support to improve their marking procedures and judgments.

Although there are several methods for introducing greater oversight and moderation into the Final Attestation, it is important to note that the success of these methods in Kazakhstan, as has been the case elsewhere, is very dependent upon the educational and cultural context of the country. Research from the United Kingdom indicated that school-based examinations were not viable because school accountability systems incentivised teachers to engage in inappropriate activities to maximise student scores on the examinations (Meadows and Black, 2018^[42]). For the Final Attestation to become more reliable and carry greater value, the national context of Kazakhstan will have to support such a transition. The policy perspectives on initial teacher education and school evaluation provide suggestions about how to create an educational environment that is more supportive and less judgmental, which will help build a context for effective and trustworthy school-based assessment.

Recommendation 3. Reinforce the educational value of the national examinations

Kazakhstan's national efforts to transform teaching and learning to be more competence-based, and assessment to be criterion-based, have not yet been reflected in the UNT, which still focuses on assessing how much information students can memorise. As a result, the examination is not helping to reinforce the curriculum or desired assessment practices and is distorting teaching and learning by narrowing what is taught and how it is assessed. On the other hand, the Final Attestation is better aligned with the curriculum, but it is not regarded seriously enough for the examination to exert significant backwash effects. This recommendation provides suggestions about how Kazakhstan can reform the development of these examinations so they can exert more positive and fewer negative backwash effects.

3.1. Establish the goal of developing fewer, but more high-quality items for the UNT

Evidence

The primary determinant of whether an examination can positively influence teaching and learning is the validity of its items. To achieve high levels of validity, national test items must be carefully created, reviewed, tested before use, and analysed and evaluated after use (American Educational Research Association, 2014^[16]; Anderson and Morgan, 2008^[43]). The goal of these rigorous procedures is not to produce a large number of items that needs to be sorted, but a smaller number of high-quality test items that assess the most important knowledge and skills that students are expected to acquire. If the process is executed properly, very few items will be eliminated after testing is complete because of appeal or because post-mortem analyses reveal improper item functioning. Box 3 describes how item validity is quality assured for the Advanced Level examinations (A-levels) in the United Kingdom.

Box 3. Quality assuring examinations for the Advanced Level examinations (A-levels) in the United Kingdom

AQA is one of the four examination boards that develop A-level examinations. These examinations are based on the knowledge, understanding and skills set out in the specification that AQA developed for each qualification. However, an examination paper only focuses on specific areas of the specification. To ensure its validity, an examination paper includes open-ended questions and complex tasks such as writing prose and constructing arguments, instead of short factual and multiple-choice questions. The examination paper is developed by an examination Committee that is composed of senior examiners and subject experts that are often teachers. Each examination Committee includes the chair of examiners for the subject tested, a chief examiner for the specification and a lead assessment writer for each examination paper.

The initial stage of developing an examination paper involves reviewing papers from previous examination series and look at how students performed at each question in order to ensure that the questions are relevant and clear, that they can be understood by students and in order to improve the quality of the examination paper.

In the second stage, the lead assessment writer develops the blueprint for the examination paper with the chief examiner. This stage is essential to the whole process from a quality assurance perspective since all the individual questions will be checked carefully against the blueprint.

The third stage consists of creating the questions and their marking scheme, reviewing them and pulling them together into the examination paper. The review of the questions is performed by the reviser and other senior examiners based on a comprehensive checklist to ensure that the test content is covered in the specification and is error free, and that each question has the right level difficulty.

The fourth stage consists of carrying out further checks to ensure the quality and consistency of the examination paper. First the qualifications developer and professional proof readers check the spelling, grammar and accuracy of the questions, as well as any images, diagrams and charts. Second, the lead assessment writer and reviser check that the paper follows the blueprint. Third the examination paper is then tested by scrutineers that are subject experts. They sit the paper in the same conditions than students and verify that all the questions are clear, have the right level of difficulty and that the paper can be completed in the allocated time. Fourth, the Question Paper Approval Committee reviews the examination paper and signs it off as fit-for-purpose and error free. Fifth, another proofreading is performed and a second scrutineer sits the paper and check it against the mark scheme. Finally, a member of the subject team who has not seen the paper before gives it a final check. If everything is fine, the chair of examiners gives their approval. Following the printing of the examination papers, a final check is made.

Source: AQA (n.d.^[44]), How a question paper is created. <https://www.aqa.org.uk/about-us/what-we-do/getting-the-right-result/how-exams-work/making-an-exam-a-guide-to-creating-a-question-paper/making-an-exam-a-guide-to-creating-a-question-paper-video-transcript> (accessed on 3 April 2020).

Low validity is the greatest concern of the UNT and is contributing to distortions in teaching and learning. The development of poor quality items, in particular those that assess memorisation instead of competences, is resulting from several aspects of the item development process. First, the overall approach to UNT item development in Kazakhstan focuses on creating a large battery of items, a small sample of which are selected each year for use. To develop an ample supply of test items, NTC selects writers from a registry of over 2 500 item developers and pays and evaluates them according to the number of items they write and review (each item is reviewed by two reviewers). These incentives create a cycle where item writers are motivated to produce more items, which increases the likelihood that poor quality items are created, which then creates the demand for even more items.

Second, regulations around item development do not effectively control for the quality of the large number of items that are produced. While a comprehensive item review process is in place, in some cases the reviewers might be item writers themselves, which is problematic because there is already concern about

the capacity of current item writers, as evidenced by the World Bank Education Modernization Project's focus on training item writers. Furthermore, owing to the large number of items, each item is not reviewed by the same reviewer, which raises concerns around the consistency of item review.

Because the World Bank's Education Modernization Project intends to train 500 item writers to develop questions that assess student competences and reflect the curriculum, this recommendation does not focus on selecting or training item developers. Instead, it focuses on developing an overall approach to item development that is based on quality, and introducing regulatory changes that can be made to support such an approach.

Recommended actions

Change the overall approach to item development

The ongoing conversations about reforming Kazakhstan's national examinations provides an opportunity not only to change the tests themselves, but also to fundamentally rethink how test development is approached in the country. The OECD recommends that NTC adopt an approach whereby the goal is to produce fewer, but higher-quality items that reflect the goals of a competence-based curriculum and criterion-based assessment. Not only would this approach improve the validity of the UNT, but, by eventually having fewer item writers, test development would be made more efficient.

From a high level perspective, these goals need to be documented in the strategic planning of the education sector. Specific goals can be set for NTC regarding how many item writers should be contracted and how many items are eliminated from the item bank each year.

Introduce external quality assurance procedures

Kazakhstan should form a diverse expert review panel whose job is to review all the items that have been developed by item writers. This type of universal review is only feasible if the item writers are creating few, high-quality items. Importantly, the panellists should not be item writers themselves. In the Kazakhstani context, it will be important that the panellists come from backgrounds that allow them to speak about the alignment of test items with the aims of the curriculum, along with having significant experience administering and developing examinations. Therefore, panellists should represent teachers, NTC, NAE and NIS. Specific activities that the panel should perform include:

- Conformance checks – to assure that items have been developed in compliance with the specifications
- Workability checks - to assure that all items and marking schemes are accurate and can be reasonably completed
- Accessibility and sensitivity checks – to assure that items are not discriminatory towards certain student groups.

Not only would these procedures directly improve item quality, but they would also indirectly improve item quality by affecting how item writers approach their task. With rigorous quality assurance procedures in place, item writers might be less focused on developing a large number of items because they understand that many would not pass review. Instead, they might concentrate on developing fewer items that are aligned with the test specifications and more likely to pass quality assurance.

3.2. Engage teachers more strongly in national testing activities to help them integrate modern assessment principles into their practice

Evidence

For examinations to exert positive backwash effects, teachers need to understand the examinations and their aims. One way of helping teachers achieve this kind of understanding is to involve them in the examination, such as by writing items and marking student responses. These procedures create a cadre of teachers who become experienced test developers and experts in assessment, who can then use what they have learned in their own practice and encourage their colleagues to do the same (OECD, 2013^[15]). Internationally, these responsibilities are often considered parts of teachers' core functions, instead of additional to them. In New Zealand, a set number of teachers are relieved of teaching responsibilities each year to administer national assessments. In Norway, marking national tests is considered professional development for teachers (OECD, 2013^[15]).

Teachers in Kazakhstan have limited interaction with the country's national examinations and thus receive little professional learning from the tests. Teachers' primary involvement with national examinations is through the Final Attestation, which they do help administer and mark, but only if they are part of their school's commission. Regarding the UNT, however, aside from the teachers who are item writers, teachers largely do not interact with it, partly owing to the confidentiality of the test.

Recommended actions

Kazakhstan should engage teachers more strongly into the country's examinations activities. Since the UNT is marked electronically, teachers cannot contribute to its marking. However, there are plans to introduce open-ended items into the UNT. As part of reviewing these items (see Recommendation 2.1.). NTC will need to determine how to mark these items. Teachers should help develop the marking schemes for these items as well as act as markers themselves. As is the case in New Zealand and Norway, these responsibilities should be incorporated into teachers' formal job expectations. As examination items rotate, there is less concern about confidentiality breaches.

Asking teachers to moderate the Final Attestation, as is suggested in Recommendation 2.2. , would also help improve teachers' assessment practices. By seeing how other teachers assess and mark the same student work, teachers can discuss the work should be assessed and, together, gain a better understanding of how to calibrate their assessment judgments vis-à-vis national learning standards. New Zealand has invested heavily in school-level assessments and has found that teacher moderation of the assessments has been important in improving assessment literacy in general (Hipkins and Robertson, 2011^[45]).

3.3. Improve the rigour and relevance of the Final Attestation

Evidence

As mentioned previously, the Final Attestation is not regarded very seriously, which prevents it from influencing the behaviour of teachers and students. This perception is partly related to the nearly universal pass rate. Nevertheless, the lack of subject-matter relevance might also be contributing to the low value of the Final Attestation. In England, example, A-levels examinations are available for over 50 subjects. This configuration helps ensure that students can demonstrate their abilities in subjects in which they are interested and which are considered valuable by the market. With only one elective combination, the Final Attestation is somewhat restrictive in what it offers. If students think the test is not relevant for them, then they will not take it seriously. Moreover, stakeholders who might be interested in a student's performance in non-core areas (e.g., employers who want to know a student's capacity in physics and

computer science) cannot learn that information through the Final Attestation, which further diminishes its value.

Kazakhstan offers different versions of some Final Attestation subjects, in particular mathematics for students who attend specialised mathematics schools. This scheme helps strengthen relevance vis-à-vis the curricula that students studied, but students are locked into which versions of the subject tests they take. A student from a non-mathematics school cannot take the more challenging mathematics test. It is likely that there are very advanced mathematics students who are not attending specialised mathematics schools. However, they cannot demonstrate their advanced understanding of the subject.

Recommended actions

Enhancing the value of an examination takes time and cannot be accomplished solely through addressing the examination itself. Associated factors, such as the needs of the economy vis-à-vis what is taught in schools, are important considerations and are outside the scope of examination reforms. Nevertheless, some examination-focused changes can be made to help achieve the aim of increasing its signalling value for all students. Some suggestions are offered below.

Review the pass rate

While the OECD does not recommend that too many students be prevented from completing upper secondary school, slightly increased rigour on the Final Attestation would help imbue it with more value. Tightening oversight measures during administration and marking (Recommendation 2.2.) might help reduce the pass rate. In addition to these measures, then NTC and the ministry can consider developing more difficult items or stricter marking schemes. If the test is more difficult to pass, then students will have more motivation to apply themselves, and teachers more motivation to help students succeed.

Increase the number of elective subjects

Many countries have reviewed their examinations systems in light of growing and diversifying upper secondary enrolment and a desire to strengthen the vocational pathway. These countries, such as North Macedonia and Serbia, have adopted a model where a set of core subjects is accompanied by several electives.

In Kazakhstan, giving students greater flexibility to choose their subjects would help to increase the relevance of the Final Attestation to them. With greater integrity and reliability, employers could also rely on Final Attestation results to help make decisions about hiring for specialised fields, which would further incentivise students to take the test seriously.

Allow all students to take different versions of subject tests

It is important that examinations assess students in core subjects to confirm that they have mastered domains (typically reading and mathematics) that are vital to their future success, and motivate teachers to help students master these domains. Nevertheless, depending on their aspirations, not all students need to demonstrate mastery of all domains to the same extent. Offering different versions of the same subject, and allowing all students to choose which version to take, enables students to show that they have learned what is required pursuant to their individual interests.

In Kazakhstan, the Final Attestation's relevance might be diminished because many students cannot choose which versions of subjects to take. Some students that are not enrolled in a specialised mathematics school may have developed an aptitude and interest in mathematics during upper secondary schooling. Likewise, some students in specialised mathematics schools might have decided that their aptitude and interests are not in mathematics. In both these cases, these students should be given an opportunity to demonstrate their full potential. This added flexibility might encourage students to view

Final Attestation as an opportunity to signal that they are indeed capable in fields for which they were not strictly trained, which might exert pressure on students and teachers to take it more seriously.

Recommendation 4. Strengthen leadership over the national examinations to ensure that they continue to fulfil their purposes effectively

Leadership over Kazakhstan’s examinations has historically been unstable. Presently, it is fragmented across several bodies and not representative of all the examinations’ stakeholders. Consequently, the examinations are not adapting successfully to changing needs, nor are decisions about them made in consideration of the purposes they should fill and their technical limitations. In order to ensure the continuation of successful examinations reforms, Kazakhstan needs to have consistent and effective leadership over the examinations.

4.1. Establish continuous leadership over Kazakhstan’s national examinations to guide their development

Evidence

Leadership over the UNT has constantly changed and has not been responsive to the changing educational landscape

Leadership responsibilities over the UNT have constantly shifted. When the UNT was first created in 2004, it was managed by the Department for Strategic Planning within the Ministry of Education. Two years later, the Committee was established and assumed management of the UNT. After the Final Attestation was developed in 2017 and the UNT only selected for entrance into tertiary education, the Division for Higher Education in the Ministry took control of the UNT. A consequence of this situation is that national examinations have not responded in an effective manner to the rapidly changing demands of the country. Many of the UNT’s limitations (e.g., having only simple multiple-choice questions that primarily assess memorisation) have been well documented in several OECD reviews since 2007. However, over a decade later, few changes have been made, partly because it is unclear who exactly is supposed to make the needed changes.

The current working group that is overseeing the UNT can help guide its immediate reforms. However, like previous leadership groups, it was formed in response to pressure to change the UNT, not to consistently manage the UNT. Continuous leadership will be necessary in order to ensure that the UNT is consistently fit-for-purpose in the context of Kazakhstan’s rapidly changing educational environment.

Leadership over the UNT and Final Attestation is separated and not representative of all stakeholders

A vital aspect of examination leadership is its composition. Bodies that make strategic decisions about an examination should represent the test’s stakeholders so a wide range of views are taken into consideration (Greaney and Kellaghan, 2007^[46]). Depending upon the examination’s purposes, a combination of the following stakeholders should help manage the tests:

- Higher education representatives
- Primary and/or secondary education representatives (students at these levels take the exams)
- Curriculum developers (the exams are based on the curriculum)

- Private and non-government agencies, whose representation is particularly important in national examinations management models to ensure that external interests are considered along with government interests.

In Kazakhstan, leadership over the UNT and Final Attestation is separated into two working groups, and neither group is representative of key stakeholders. The working group that is leading the UNT primarily represents higher education but not secondary education, which is the level where students take the UNT. The working group that is leading the Final Attestation primarily represents secondary education, but not curriculum developers, though the test is based on the curriculum.

Having leadership with limited representation risks that not all purposes and important characteristics of the test are not considered in the upcoming reforms. For example, without secondary education represented in the UNT working group, there is little assurance that the decisions about the test will try to lessen its distortive effects on teaching and learning in secondary education classrooms. Furthermore, there is limited scope to think about the complementarity of the two tests, which risks that inefficiencies and overlaps will develop.

Recommended actions

As a point of principle, leadership over examinations should be consistent, expert and representative. A common method of building such leadership is to create steering committees to lead examinations development. In the Netherlands, such a committee led development of an examination following the introduction of a new academic subject (Michels and Eijkelhof, 2019^[47]).

The policy perspective on national assessments recommends that Kazakhstan establish a steering committee to lead the development of the EAAA. This policy perspective similarly recommends that Kazakhstan establish a steering committee to lead the development of the UNT and Final Attestation. This committee will be responsible for setting the overall, long-term strategy of the examinations. As the strategy is implemented, the steering committee would make necessary adjustments to the examinations in consideration of national priorities and informed by data generated by the tests.

Since Kazakhstan already has two separate working groups leading the Final Attestation and UNT, representatives from those two groups, in particular from the Divisions of Primary and Secondary and Higher Education, should sit on the new steering committee. The NIS Centre for pedagogical Measurement and NTC (see Recommendation 4.2.) should also be invited to contribute their technical expertise. University leadership and private sector representatives should also join the committee, as these two sectors have vested interests in the signalling value of the examinations. This configuration would establish continuous, effective leadership over the examinations and, because the same organisation is responsible for all of them, would enable decisions to be made in consideration of the complementarity of the tests.

4.2. Empower NTC to contribute to decision-making over national examinations and assessments

Evidence

In most mature education systems, examinations authorities are responsible for executing the vision of the country's national examinations, as expressed by the steering committee. Nevertheless, because they create and administer the tests and are measurement specialists, examinations authorities are well positioned to advise the steering committee on the strategic direction of the examinations. Often, the examinations authority sits on the steering committee, which gives it the remit to contribute to the development of the examinations. This arrangement is constructive because other members of the steering committee bring representative leadership to the table, but the examinations authority brings familiarity with the examinations and technical expertise.

In Kazakhstan, NTC's mandate is primarily to execute the plans of other bodies who do have leadership responsibilities, such as the Committee. NTC itself has a minimal a leadership role and there is limited space for NTC to provide feedback to examination leadership based upon its experiences. It is not represented in the working groups that are leading the UNT or the Final Attestation. Consequently, the decisions that are made about the examinations might be well-intentioned, but sometimes problematic from a technical measurement perspective.

Recommended actions

Invite NTC to sit on the steering committees of the national examinations and the EAAA

There are several reasons why NTC has a limited role in leading examinations. One is its formal exclusion from key policy-making processes about examinations. While NTC's remit is to act as an executor of policy decisions, its remit should not preclude it from contributing to making those decisions. This recommendation, and the policy perspective on the national assessment, have suggested that steering committees be created to lead the development of Kazakhstan's national examinations and the EAAA. The president of NTC should be invited to sit on both those committees, which will allow NTC's technical expertise to help direct the developments of the tests.

Elevate the status of NTC by strengthening its governance structure

For NTC to help steer Kazakhstan's examinations effectively, it needs significant public standing so its advice carries weight. Internationally, examinations authorities often derive their public standing through the composition of their leadership. Rather than being led by an individual, they are typically led by management boards comprised of the chief executive of the examination authority, government leadership (not exclusive to the Ministry of Education) and prominent representatives from the private and non-profit sectors. These boards are responsible for overseeing the activities of the examinations authorities and aligning them with the public interest. Having strong, influential management boards imbues examinations authorities with credibility and externality, which elevates their standing and enables them to provide credible advice. Box 4 discusses the how the examinations agency of Hong Kong is governed.

Box 4. Governance of the Hong Kong Examinations and Assessment Authority

The Hong Kong Examinations and Assessment Authority (HKEAA) is a statutory body that oversees and administers all public examinations and assessments in Hong Kong. These include the Hong Kong Diploma of Secondary Education, which acts as a secondary school leaving and tertiary entrance examination, and the Territory-wide System Assessment, a system evaluation tool.

HKEAA is led by a Secretary General and the organisation is managed by a management board, called the Council. Seventeen individuals sit on the Council, including the Secretary General. Other members include representatives from the education sector, universities, other government bodies and private industry leaders. The Council is responsible for formulating examination policies and monitoring the work of HKEAA.

The Council appoints several standing committees to advise its decision-making and support the major activities of HKEAA. Several working groups and sub-committees can also be convened regarding specific issues. Standing committees include an information technology committee, a standards setting committee, a committee on the special needs testers and numerous subject (academic domain) committees.

Source: (HKEAA, 2020^[48]). Hong Kong Examinations and Assessment Authority – Governance. http://www.hkeaa.edu.hk/en/about_hkeaa/governance/ (accessed on 8 April 2020).

Currently, NTC leadership reports directly to the Ministry of Education, which makes it difficult for NTC to provide constructive feedback regarding ministry decisions. The OECD recommends that NTC be led by a management board, on which the NTC president would sit. NTC's management board should represent the diverse stakeholders of the country's examinations, including the Ministry of Education, universities, other government entities and the private sector. The OECD suggests that Kazakhstan consider the representatives listed in Table 5 to serve on NTC's management board.

Table 5. Suggested composition of the NTC management board

Suggested representative	Purpose
Senior Ministry of Education leadership	Represent national education interests
Committee for Quality Assurance	Ensure NTC's compliance with regulatory measures
University leadership	Represent the interests of the higher education sector
Private-sector leadership	Represent the needs of employers and non-government interests
NTC leadership	Represent the opinions of NTC staff

4.3. Develop technical reports about the tests and use them to make future decisions about test development procedures and design

Evidence

Given the importance of national assessments and examinations and the highly technical nature of the tests, it is imperative that decisions over their operations and designs be based on a rigorous study of the evidence that is generated by test administrations. In many countries, national testing centres produce detailed, technical reports about the tests that they oversee. These reports contain information about item characteristics from pilot testing and post-mortem analyses, how long students spend on each item, and which items received appeals from students. The reports also contain overall meta-analyses about the tests, such as the capacity of one subject to predict success on another.

In Italy, the national education system evaluation institute (INVALSI) publishes technical reports detailing how the yearly national assessments were designed and developed. These reports include pre-testing information as well as the range of statistics used to exclude poor quality items (INVALSI, 2019^[49]). In England, tests are provided by different examination boards, so each one analyses item characteristics separately and independently. A wealth of information is available regarding the maintenance of common standards across subjects. For example, there are yearly reports on the quality of marking, reviews of marking and moderation showing the number of appeals managed by the examination system, and centralised research about the comparability of results between different subjects and subject areas.

Importantly, these reports are used to improve test development procedures and test designs. If, for example, analyses consistently reveal that items are too difficult, then test developers should be asked to write less difficult items and be shown examples of such items as identified by the reports. At a higher-level, if student results on several items are highly correlated, then consideration can be given to eliminating some of the items since it would not alter the reliability of the final results.

Post-mortem analyses of national tests in Kazakhstan occur in a very structured manner. Students are given a window of time during which they can appeal and NTC procedures specify that item characteristics be analysed statistically. Nevertheless, there does not exist annual, comprehensive, technical reporting for the national tests that detail the results of these activities. Furthermore, technical information produced by post-mortem analyses do not seem to strongly shape test development. Nor do high-quality items, again identified by analyses, act as exemplars in the materials that are given to test developers and reviewers. The development of the tests, therefore, continues in an ad hoc manner and is largely uninformed by empirical evidence.

Recommended actions

NTC should annually develop technical reports of the operations and functioning of the tests that they oversee. These reports should form the basis of NTC's (specifically its management board) decision-making regarding the tests. In this policy perspective, the OECD has not made specific recommendations regarding exactly how many subjects should be assessed or how difficult items should be. Instead, the OECD has provided general guidance regarding these decisions and suggests that they, and others, be made only after sufficient information has been produced to inform the decision-making process.

One important issue that should be consistently reviewed, based on technical evidence, is the length of examinations. Post-mortem analyses will show the difficulty and discrimination levels of all items. Statistical models can then be produced to determine what the optimum number of items to include for each subject in consideration of producing reliable results but alleviating testing burden. For example, if removing the ten weakest items from the language test of the UNT would produce student results that are not statistically significantly different, then those items can be removed to reduce testing burden without impacting the rigour of the results.

4.4. Begin the development a holistic strategy regarding the national examinations that considers broad factors that influence their effectiveness

Evidence

When Kazakhstan has reformed its examinations, the focus has typically been to change the examinations system itself. These reforms, however, will not be able to achieve their goals unless the wider educational environment is also considered. For example, Kazakhstan has made concerted efforts to reduce the negative backwash effects generated by the UNT, such as making the test more flexible and creating the Final Attestation. . Despite these efforts, however, the pressure generated by the UNT has not diminished and private tutoring is still very common. The reason this situation has persisted is not solely related to the

examination itself. It is also determined by the economic competitiveness conferred by a higher education qualification and the fact that public scholarships to attend university are limited (OECD, 2017^[11]).

Recommended actions

With a strongly governed NTC managing its national examinations, Kazakhstan will have the leadership needed to develop a holistic strategy around its national examinations. This strategy, in addition to addressing the features of the examinations, should consider broad factors that impact the success of examinations reform. Specific topics that the OECD recommends be included in the strategy are:

Strengthening vocational and technical pathways

The demand for university education in Kazakhstan is straining the higher education sector, which struggles to keep up with demand while maintaining quality. In this context, entrance examinations become more important because more students are competing for fewer spaces, and even fewer spaces at prestigious institutions (OECD, 2017^[11]). Strengthening alternative pathways, which has been suggested in depth in previous OECD reviews (OECD, 2014^[9]), would provide job seekers with a viable alternative to entering university, which would decrease the competitive pressure around the UNT. An important issue is that, despite the fact that almost 40% of upper secondary students attend vocational colleges, there is no nationally standardised way of certifying their competences upon completion, which can negatively affect their employment potential. Specific steps Kazakhstan can consider to address this issue include:

- Extending the Final Attestation to certify that vocational students have mastered core domains through different versions of certain subject tests, which would also motivate colleges and teachers in colleges to focus on developing their students' foundational academic competences.
- Incorporating vocational subjects and aptitudes into the Final Attestation as electives so vocational students can have national certification of their abilities
- Further development and finalisation of the national qualification framework

Improving access to university

Related to the importance of a university degree is access to higher education. University scholarships become more precious if financial barriers make enrolment into universities difficult. Previous OECD reviews recommended that Kazakhstan consider establishing student loan schemes alongside scholarships to improve access to higher education. More funding opportunities would reduce the pressure students face when taking the UNT (sometimes repeatedly) in order to attain a high enough score to qualify for a scholarship. Acting upon these recommendations will be important not just to improve in access to higher education, but also to reduce the negative backwash the UNT exerts on students in upper secondary education.

Continue reviewing the dual-examinations system in the long-term

Previous OECD reviews recommended that Kazakhstan develop a dual-examinations system at the end of upper secondary education. This recommendation was made in consideration of the Kazakhstani context at the time, which had a single, very high-stakes examination and wanted to assess higher-order skills while reducing the negative backwash effects of the single examination. Given the noted challenges of the UNT, the OECD recognised that expanding the examination's functions (i.e., making it also certify completion of upper secondary education) would not accomplish these goals. In fact, adding to the purposes of the UNT would likely make the UNT even more high-stakes while still assessing students using fact-based items. Developing a separate examination was thus recommended as it represented a clear separation from the legacy of UNT.

As the Kazakhstani education system matures and the issues mentioned above become less pressing, then Kazakhstan can continue to review its examinations at the end of upper secondary education. The introduction of the Final Attestation was a very positive development and more can be done to strengthen it. Reinforcing the value of an examination means improving its integrity and reliability, which, in almost all countries, implies that the test be administered externally. When Kazakhstan is ready to consider external administration of the Final Attestation, it will also need to consider whether to maintain a dual-examinations system, or transition to a single examination system, as many OECD countries have done, in which one examination serves to certify completion of upper secondary education and select students for tertiary education. A single examination, based on the curriculum, can be more equitable because it diminishes the need for students to seek out private tutoring. In a country as vast as Kazakhstan, managing and administering only one examination would also be considerably more efficient, and less burdensome on students, than two.

Using one examination to both certify and select students is very challenging. The same test must allow students to demonstrate basic minimum competences and advanced understanding, while maintaining high levels of integrity, reliability and validity. Many of the methods that other countries employ to achieve this balance, such as allowing students to choose elective subjects and offering different versions of subjects, already appear to some degree in Kazakhstan's dual examinations. As these features are refined and the examinations' characteristics continuously strengthened, NTC can regularly examine the examinations' technical reports to make a determination about if a switch to a single examination model would be advisable.

References

- American Educational Research Association (2014), *Standards for Educational and Psychological Testing*. [16]
- Anderson, P. and G. Morgan (2008), *Developing Tests and Questionnaires for a National Assessment of Educational Achievement*, The World Bank, Washington, D.C. [43]
- AQA (n.d.), *How a question paper is created*, <https://www.aqa.org.uk/about-us/what-we-do/getting-the-right-result/how-exams-work/making-an-exam-a-guide-to-creating-a-question-paper/making-an-exam-a-guide-to-creating-a-question-paper-video-transcript> (accessed on 3 April 2020). [44]
- Asian Development Bank (2017), “Innovative Strategies for Accelerated Human Resource Development in South Asia: Student Assessment and Examination—Special Focus on Bangladesh, Nepal, and Sri Lanka”, <http://dx.doi.org/10.22617/TCS179079>. [34]
- Baartman, L. et al. (2007), “Evaluating assessment quality in competence-based education: A qualitative comparison of two frameworks”, *Educational Research Review*, Vol. 2/2, pp. 114-129, <http://dx.doi.org/10.1016/j.edurev.2007.06.001>. [8]
- Baird, J. et al. (2018), *Examination standards: How measures and meanings differ around the world*. [40]
- Bishop, J. (1999), *Are National Exit Examinations Important for Educational Efficiency?*, Cornell University ILR School, <http://digitalcommons.ilr.cornell.edu/articles> (accessed on 10 January 2019). [6]
- Bray, M. and M. Kobakhidze (2014), “The Global Spread of Shadow Education”, http://dx.doi.org/10.1007/978-94-6209-650-9_11. [22]
- Bray, M. and O. Kwo (2013), “Behind the façade of fee-free education: shadow education and its implications for social justice”, *Oxford Review of Education*, Vol. 39/4, pp. 480-497, <http://dx.doi.org/10.1080/03054985.2013.821852>. [21]
- Chamberlain, S. (2013), “Communication strategies for enhancing qualification users’ understanding of educational assessment: Recommendations from other public interest fields”, *Oxford Review of Education*, Vol. 39/1, pp. 114-127, <http://dx.doi.org/10.1080/03054985.2013.764757>. [38]
- Department Basic Education Republic of South Africa (2018), *Annual performance plan 2018/2019*, <https://www.education.gov.za/Portals/0/Documents/Reports/Annual%20Performance%20Plan%20201819.pdf?ver=2018-03-14-121624-263> (accessed on 3 April 2020). [36]
- ETS (2019), *Content and Construct Validity*, https://www.ets.org/gre/institutions/about/fairness/content_construct_validity/ (accessed on 26 March 2020). [17]
- Government Communication and Information Systems Republic of South Africa (2018), *Official guide to South African 2017/2018 Education*, [35]

- <https://www.gcis.gov.za/sites/default/files/docs/resourcecentre/pocketguide/08-Education-1718.pdf> (accessed on 3 April 2020).
- Greaney, V. and T. Kellaghan (2007), *National Assessments of Educational Achievement Volume 1*, The World Bank, <http://dx.doi.org/10.1596/978-0-8213-7258-6>. [46]
- Hipkins, R. and S. Robertson (2011), *Moderation and Teacher Learning What can research tell us about their interrelationships?*, <https://www.nzcer.org.nz/system/files/moderation-teacher-learning.pdf> (accessed on 1 April 2020). [45]
- HKEAA (2020), *Hong Kong Examinations and Assessment Authority - Governance*, http://www.hkeaa.edu.hk/en/about_hkeaa/governance/ (accessed on 8 April 2020). [48]
- IAC (2019), *National Report on the State and Development of the Education System of the Republic of Kazakhstan 2018*, Information Analytic Centre, Nur-Sultan. [4]
- INVALSI (2019), *National Institute for the Evaluation of the Education and Training System*, <https://www.invalsi.it/invalsi/index.php> (accessed on 25 February 2020). [49]
- Japan Times (2019), *Open-ended questions for Japan's new university entrance exams scrapped*, https://www.japantimes.co.jp/news/2019/12/17/national/open-ended-questions-scrapped-university-entry-exams-japan/#.Xim_QMhKiHs (accessed on 23 January 2020). [27]
- Jumabayeva, Z. (2016), "The Key Drivers of the Unified National Test in Kazakhstan: A Critical Analysis of its Impact on School Leavers", *NUGSE Research in Education*, Vol. 1, https://www.researchgate.net/publication/311665473_The_Key_Drivers_of_the_Unified_National_Test_in_Kazakhstan_A_Critical_Analysis_of_its_Impact_on_School_Leavers (accessed on 27 January 2020). [33]
- Kellaghan, T. and V. Greaney (2019), *Public Examinations Examined*, The World Bank, <http://dx.doi.org/10.1596/978-1-4648-1418-1>. [14]
- Kirkpatrick, R. (2011), *The Negative Influences of Exam-Oriented Education on Chinese High School Students: Backwash from Classroom to Child*, <https://link.springer.com/content/pdf/10.1186/2229-0443-1-3-36.pdf> (accessed on 30 March 2020). [20]
- Kitchen, H. et al. (2019), *OECD Reviews of Evaluation and Assessment in Education: Student Assessment in Turkey*, OECD Reviews of Evaluation and Assessment in Education, OECD Publishing, Paris, <https://dx.doi.org/10.1787/5edc0abe-en>. [7]
- Kitchen, H. et al. (2017), *Romania 2017*, OECD Reviews of Evaluation and Assessment in Education, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264274051-en>. [41]
- Kitchen, H. et al. (2019), *OECD Reviews of Evaluation and Assessment in Education: North Macedonia*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/079fe34c-en>. [24]
- Leiva, F., F. Ríos and T. Martínez (2006), "Assessment of interjudge reliability in the open-ended questions coding process", *Quality and Quantity*, Vol. 40/4, pp. 519-537, <http://dx.doi.org/10.1007/s11135-005-1093-6>. [26]
- Madaus, G. (1988), "The Distortion of Teaching and Testing: High-Stakes Testing and Instruction", *Peabody Journal of Education*, Vol. 65/3, pp. 29-46, <http://dx.doi.org/10.1080/01619568809538611>. [19]

- Meadows, M. and B. Black (2018), “Teachers’ experience of and attitudes toward activities to maximise qualification results in England”, *Oxford Review of Education*, Vol. 44/5, pp. 563-580, <http://dx.doi.org/10.1080/03054985.2018.1500355>. [42]
- Michels, B. and H. Eijkelhof (2019), “The Development and Implementation of an Interdisciplinary STEM Course for Upper Secondary Schools”, *Educational Designer*, Vol. 2/3, <https://www.educationaldesigner.org/ed/volume3/issue12/article48/> (accessed on 6 May 2020). [47]
- New Zealand Qualifications Authority (n.d.), *External Advisory Groups*, <https://www.nzqa.govt.nz/about-us/our-structure/advisory-groups/> (accessed on 6 April 2020). [29]
- OECD (2019), *PISA 2018 Results (Volume I): What Students Know and Can Do*, PISA, OECD Publishing, Paris, <https://dx.doi.org/10.1787/5f07c754-en>. [3]
- OECD (2018), “Education Policy Outlook Kazakhstan”, <https://www.oecd.org/education/Education-Policy-Outlook-Country-Profile-Kazakhstan-2018.pdf>. [10]
- OECD (2017), *Education in Lithuania*, Reviews of National Policies for Education, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264281486-en>. [12]
- OECD (2017), *Higher Education in Kazakhstan 2017*, Reviews of National Policies for Education, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264268531-en>. [11]
- OECD (2015), *Education at a Glance 2015: OECD Indicators*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/eag-2015-en>. [13]
- OECD (2014), *Reviews of National Policies for Education: Secondary Education in Kazakhstan*, Reviews of National Policies for Education, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264205208-en>. [9]
- OECD (2013), *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*, OECD Reviews of Evaluation and Assessment in Education, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264190658-en>. [15]
- OECD (2007), *Reviews of National Policies for Education: Higher Education in Kazakhstan 2007*, Reviews of National Policies for Education, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264033177-en>. [32]
- OECD/The World Bank (2015), *OECD Reviews of School Resources: Kazakhstan 2015*, OECD Reviews of School Resources, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264245891-en>. [5]
- OECD/The World Bank (2007), *Reviews of National Policies for Education: Higher Education in Kazakhstan 2007*, Reviews of National Policies for Education, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264033177-en>. [31]
- Opposs, D. et al. (2020), “Governance structure and standard setting in educational assessment”, *Assessment in Education: Principles, Policy & Practice*, pp. 1-23, <http://dx.doi.org/10.1080/0969594X.2020.1730766>. [28]
- Pearson (2018), *PARCC Final Technical Report for 2018 Administration*, <https://files.eric.ed.gov/fulltext/ED599198.pdf> (accessed on 31 March 2020). [37]
- Republic of Kazakhstan (2016), “State Program for the Development of Education 2016-19”, 205. [1]

- Rostiashvili, K. (2004), *Corruption in the Higher Education System of Georgia*, [18
] https://tracc.gmu.edu/sites/default/files/Rostiashvili_Corruption_in_HES_Eng.pdf (accessed on 30 March 2020).
- Simpson, L. and J. Baird (2013), "Perceptions of trust in public examinations", *Oxford Review of Education*, Vol. 39/1, pp. 17-35, <http://dx.doi.org/10.1080/03054985.2012.760264>. [39
]
- Tanujaya, B., J. Mumu and G. Margono (2017), "The Relationship between Higher Order Thinking Skills and Academic Performance of Student in Mathematics Instruction, International Education Studies, 2017", *International Education Studies*, Vol. 10/11, [25
] <https://eric.ed.gov/?id=EJ1159551> (accessed on 25 February 2020).
- Tong, C., C. Lee and G. Luo (2020), "Assessment reform in Hong Kong: developing the HKDSE to align with the new academic structure", *Assessment in Education: Principles, Policy & Practice*, pp. 1-17, <http://dx.doi.org/10.1080/0969594X.2020.1732866>. [30
]
- World Bank (2018), *Population ages 0-14 (% of total population) - Kazakhstan, OECD members | Data*, <https://data.worldbank.org/indicator/SP.POP.0014.TO.ZS?locations=KZ-OE> (accessed on 3 March 2020). [2]
- Zhang, Y. (2013), "Does private tutoring improve students' National College Entrance Exam performance?-A case study from Jinan, China", *Economics of Education Review*, Vol. 32/1, pp. 1-28, <http://dx.doi.org/10.1016/j.econedurev.2012.09.008>. [23
]

Annex A. Key indicators

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

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

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

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

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

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

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

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

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

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