



Key Policy Areas for Student Assessment

Status

1. Classroom Assessment

Classroom assessment in Belarus is supported by formal, system-level documents that provide comprehensive information on what students are expected to learn and how classroom assessment should be conducted by teachers. Mechanisms are in place to monitor the quality of classroom assessment practices, and teachers engage in positive assessment practices in the classroom. At the same time, there is a need to improve the quality and availability of resources that support teachers' classroom assessment practices and to ensure that high-quality training opportunities are available to teachers on a regular basis.



2. Examinations

Centralized Testing is an examination program that is used to select students into higher and specialized secondary education in Belarus. The program is overseen by the Republic Institute for Knowledge Control (RIKC), is adequately staffed, and is well resourced through government funding. The examination is well aligned with other assessment activities in the country, and extensive quality assurance procedures are used to help ensure its technical quality. At the same time, there is a need to better ensure equitable access to the examination, including by students with disabilities, and to introduce mechanisms to monitor the impact of the examination on the education system and on students.



3. National Large-Scale Assessment (NLSA)

Belarus's National Large-Scale Assessment exercise—the Republican Monitoring of Education Quality (“Republican Monitoring”)—has been administered annually since 2003. The Republican Monitoring exercise aims to assess the degree to which students have acquired the content taught under the formal educational program. The Republican Monitoring exercise is fully standardized at the system level, and sufficient funding is provided through national and local budgets to cover all core assessment activities. The Department of Education Quality Monitoring, which oversees the assessment, has most of the necessary resources to carry out its responsibilities, although there is room for improvement in the availability of computers for technical staff, software, and servers.



4. International Large-Scale Assessment (ILSA)

Belarus has never participated in an ILSA but is taking active steps to participate in the 2018 administration of the Program for International Student Assessment (PISA). RIKC will be responsible for PISA 2018 activities in the country. RIKC has extensive experience carrying out assessments within Belarus. RIKC specialists, as well as staff from the National Institute of Education who will help with PISA 2018 activities, are well prepared to commence work on the ILSA. To maximize the benefits of participating in PISA 2018, the country may wish to consider doing a detailed needs assessment and developing a plan to address areas where they will need particular technical support.



Table of Contents

Introduction	3
What Is SABER–Student Assessment?	3
Education in Belarus.....	5
Classroom Assessment in Belarus	7
Examinations in Belarus	9
National Large-Scale Assessment in Belarus.....	12
International Large-Scale Assessment in Belarus	16
Appendix 1: Assessment Types and Their Key Differences.....	18
Appendix 2: Summary of the Development Levels for Each Assessment Type	19
Appendix 3: Methodology for Assigning Development Levels	20
Appendix 4: Stakeholders Consulted during Completion of the SABER–Student Assessment Questionnaires in Belarus	21
Appendix 5: SABER–Student Assessment Rubrics for Belarus	23
Acknowledgments.....	66
References.....	66

Introduction

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

What Is SABER–Student Assessment?

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

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

1. Providing information on levels of student learning and achievement in the system
2. Monitoring trends in education quality over time
3. Supporting educators and students with real-time information to improve teaching and learning and
4. Holding stakeholders accountable for results.

SABER–Student Assessment methodology

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

Assessment Types and Purposes

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

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

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

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

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

Quality Drivers of an Assessment System

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

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

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

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

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

Table 1: Framework for Building an Effective Assessment System, with Indicator Areas

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

Source: World Bank.

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

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

Levels of Development

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

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

Rubrics are used to evaluate data collected using the standardized questionnaires. The goal of the rubrics is to provide a country with some sense of the development level of its assessment activities compared to best or recommended practice. For each indicator, the rubric displays four development levels: *Latent*, *Emerging*, *Established*, and *Advanced*. These levels are artificially constructed categories chosen to represent key stages on the underlying continuum for

each indicator. Each level is accompanied by a description of what performance on the indicator looks like at that level.

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

A summary of the development levels for each assessment type is presented in appendix 2.

In reality, assessment systems are likely to be at different levels of development in different areas. For example, a system may be *Established* in the area of examinations, but *Emerging* in the area of large-scale, system-level assessment, and vice versa. Although intuition suggests that it is probably better to be further along in as many areas as possible, the evidence is unclear as to whether it is necessary to be functioning at *Advanced* levels in all areas. Therefore, one might view the *Established* level as a desirable minimum outcome to achieve in all areas but aspire beyond that only in those areas that most contribute to the national vision or priorities for education. In line with these considerations, the ratings generated by the rubrics are not meant to be additive across assessment types (that is, they are not meant to be added to create an overall rating for an assessment system; they are meant only to produce an overall rating for each assessment type). The methodology for assigning development levels is summarized in appendix 3.

Education in Belarus

Belarus is an upper-middle income country in Eastern Europe. GDP per capita (current \$) is \$8,040. Between 2001 and 2008, GDP growth averaged 8.3 percent annually. However, growth has slowed considerably since then, to approximately 1.6 percent in 2014, in part because of a reduction in export demand and access to external borrowing.

Although preschool education is not compulsory in Belarus, 99 percent of children attend early childhood development programs before starting school at the

age of six. In accordance with statute 2 of the Code of the Republic of Belarus on education in the Republic of Belarus, general basic education is mandatory, and this includes elementary education (grades 1–4) and basic education (grades 5–9). Upon completion of grade 9 (generally at age 15), students must pass exit examinations to receive a certificate of completion of basic education. Students can continue their secondary education in institutions of general secondary education (grades 10–11), institutions of professional-technical education, or institutions of secondary-specialized education. Students who continue through general secondary education and complete grade 11 must pass exit examinations to receive a certificate of general secondary school completion. In May 2015, Belarus joined the Bologna Process as a way to further enhance its higher education system and align it with international standards.

The education system in Belarus faces several challenges despite having made major progress. Since its independence in 1991, Belarus has expanded access to education at the preprimary and secondary levels. However, slowed economic growth and an aging and declining population have increased attention to the need for the country to increase efficiency within the education system and workers' productivity. In response to these needs, efforts to improve education quality and optimize the education system have been introduced. One such initiative has focused on increasing financial efficiencies within the system as schools are consolidated and students are transported from areas with a smaller student population to areas with a larger student population under the National Development Program for General Secondary Education (2007–16). Although the school consolidation program has increased efficiencies, it has not been undertaken in parallel with quality improvements. For example, schools that have remained open and are receiving new students from areas with lower populations have not been provided with additional quality-enhancing inputs, such as improved learning environments, learning materials, or modern infrastructure, equipment, and technology.

In response to these and other challenges, Belarus is prioritizing the reform of learning environments to enhance quality by improving schools' infrastructure and expanding access to laboratories and information technology. In addition, to improve the quality of education, Belarus is interested in improving its assessment system to have better data on student

performance that can inform further education reforms and policies.

Detailed information was collected on Belarus's student assessment system using the SABER–Student Assessment questionnaires and rubrics to benchmark it against best practices. Specifically, a local consultant with in-depth knowledge of, and experience with, the education system in Belarus oversaw the completion of the four SABER–Student Assessment questionnaires: one questionnaire each for Classroom Assessment, Examinations, National Large-Scale Assessment, and International Large-Scale Assessment. The data to complete these questionnaires were obtained through interviews and focus groups with key stakeholders¹ and a review of existing official and technical documents. The information in the completed questionnaires was then applied to the SABER–Student Assessment rubrics (one rubric for each assessment type), and the conclusions of this report were determined on the basis of this analysis. It is important to remember that these tools primarily focus on benchmarking a country's policies and arrangements for assessment activities at the system or macro level. Additional data would need to be collected to determine actual, on-the-ground practices in Belarus, particularly by teachers and students in schools.

The following sections discuss the findings by each assessment type, accompanied by suggested policy options. The suggested policy options were determined in collaboration with key local stakeholders based on Belarus's immediate interests and needs. Detailed, completed rubrics for each assessment type are provided in appendix 5.

¹ The list of stakeholders consulted is provided in appendix 4.

Classroom Assessment in Belarus

Level of Development: ESTABLISHED

Several formal, system-level documents provide guidelines for classroom assessment in Belarus. For example, the documents “Educational Standards for School Subjects, 2009” and “Educational Programs for Extracurricular Subjects, 2009–2013” offer useful guidance for classroom assessment activities by specifying what students are expected to learn in different subjects at different grade levels. These documents do not, however, describe the level of performance that students are expected to reach in each subject area (that is, how well students should learn something). Another official document, “Educational Standards for School Subjects” (2012, 2015), not only outlines what students are expected to learn, but also specifies the level of performance that they should demonstrate in each subject area. All of these documents are publicly available.

Other official and publicly available documents provide more specific guidance for classroom assessment activities in general secondary education, including describing what should be assessed, criteria for scoring student work, formats and methods for carrying out classroom assessment activities, mechanisms for creating scores and marks, and record keeping.²

Most teachers use classroom assessment information to diagnose student learning issues, provide continuous feedback to students on their learning, plan further instruction, and evaluate student performance. Additionally, classroom assessment information is required to be used for certification of student achievement and as an input for selection to the next educational level; guidance for these uses is provided in official documentation made available to schools. According to the data provided in the completed SABER–Student Assessment questionnaire, a marginal number of teachers in Belarus engage in poor classroom assessment practices, with some teachers engaging in grade inflation and overemphasizing the assessment of memorization/recall of information.

² These documents include “Assessment Norms for Outcomes of Educational Activity of Students in Educational Subjects,” “On Organization of Work of General Secondary Education Institutions to Monitor and Evaluate the Outcomes of Learning Activity of Students during the

Schools report individual student performance on classroom assessment in all subject areas to students and their parents through schoolwide and class-specific parent-teacher conferences. Student grades are recorded in a classroom journal and in the student’s grade journal (*dnevnik*). Parents are required to review the student’s grade journal every week and every quarter. There is currently a move toward maintaining the classroom journal and the student’s grade journal in electronic form. Parents can also receive information on their child’s performance in the classroom through SMS.

According to the data provided in the completed SABER–Student Assessment questionnaire, extensive resources are available to teachers for their classroom assessment activities and for developing competency in classroom assessment. All teachers receive textbooks that contain questions and sample material for conducting classroom assessment activities. Compilations of materials to be used for conducting continuous, intermediate, and final assessment have been published for all school subjects in grades 1–11 and are available for purchase. All subject teachers receive funds specifically to acquire the methodological literature. Diagnostic assessment modules for all subjects in grades 1–11 are available on the national education portal. All participants in the education process, including students, parents, and teachers, have access to them. Although the sample questions for classroom assessment activities and computer-based resources are considered to be of high quality, textbooks are considered to be of medium quality. In addition, although preservice teacher training in classroom assessment is available to most teachers, it is neither of high quality nor available on an annual basis. Teacher self-education opportunities are available to all teachers. For example, in institutions of general secondary education teachers independently select topics for self-education. Specifically, teachers review the educational, pedagogical, psychological, and methodological literature related to their chose topic; develop lessons on their selected topic; present and discuss the topic at methodological meetings that

Ungraded Learning Stage I of General Secondary Education,” and “Rules for Certification of Pupils at Mastering Content Educational Programs of General Secondary Education.”

bring together subject teachers; and publish their methodological findings in methodological subject journals, in the newspaper *Nastaunitskaya Gazeta*, on the websites of their education institutions, and on their personal websites. Teachers present their publications on and experience with their chosen self-education topic during an examination that they take to confirm their existing level of teaching qualification or to receive the next level of teaching qualification. Teachers also benefit from groups that are formed within the education institutions that provide a forum for teachers to discuss various topics in education, such as methodological approaches in various subject areas and assessment of student learning.

A number of systems are in place to monitor the quality of classroom assessment practices, including as a part of school inspection and teacher supervision and performance evaluation. In addition, high-level reviews of the quality of education in Belarus include a focus on classroom assessment, and government funding is awarded to researchers and practitioners for research and development on classroom assessment practices. An external moderation system consists of subject-area commissions at the district level that objectively evaluate teachers' review of examination questions of students who are being considered to receive an attestation of general secondary education with an award of a gold (or silver) medal. At the same time, a more wide-reaching moderation system, or a moderation system where teachers from different schools meet to review assessment tasks, scoring criteria, and students' work, with the purpose of ensuring the validity and comparability of student scores and grades, is not in place.

Suggested policy options:

1. **Increase the quality and availability of resources that support teachers' classroom assessment practices. For example:**
 - a. Improve the quality of student textbooks that provide support for classroom assessment.
 2. **Ensure that high-quality training opportunities are available to all teachers on a regular basis. For example:**
 - a. Review preservice teacher training options for building competencies in classroom assessment to ensure they are of high quality and available on an annual basis.
- b. Institute annual in-service teacher training opportunities on classroom assessment (including via live and prerecorded courses delivered via computer) to ensure that teachers have the ability to hone their classroom assessment knowledge and skills every year as needed.
 - c. Ensure that distance learning options are available to all teachers.

Examinations in Belarus

Level of Development: ESTABLISHED

Results from Belarus’s national examination program, “Centralized Testing,” are used to select individuals into higher and specialized secondary education institutions. Centralized Testing can be taken by those individuals who have completed the educational program of general secondary education, of the professional-technical education that provides vocational qualification and general secondary education, or of specialized secondary education as confirmed by a document of completion of such education. Results are also used to inform policy and evaluate interventions aimed at improving student learning. The Council of Ministers of the Republic of Belarus Resolution no. 714, “On Approval of the Regulation on the Procedure for Organizing and Holding Centralized Testing” (June 6, 2006), is the formal, publicly available policy document that authorizes the examination program. Guidance on various aspects of Centralized Testing is provided in Ministerial Orders and Resolutions. For example, the guidance on the rules of computing examination scores and the organizational rules at the centers carrying out the Centralized Testing is provided within the Order of the Ministry of Education of the Republic of Belarus Resolution no. 323 of April 22, 2015.

The Republican Institute for Knowledge Control (RIKC)—accountable to the Ministry of Education, from which it is institutionally separate—has been responsible for many elements of the examination since 2000.

The Centralized Testing program is well aligned with classroom assessment and National Large-Scale Assessment activities in the country in terms of the content and skills being measured. The Centralized Testing program measures student achievement in relation to the official learning standards and curriculum in Belarus. This focus is ensured through officially mandated internal and external alignment reviews that take place in all examination rounds.

Overall, the Centralized Testing program is well resourced and is sufficiently funded by the government. Staff possess relevant qualifications for their jobs, and no major issues are seen with their

effectiveness. Classroom teachers do not perform any examination-related tasks.

Opportunities are available to interested stakeholders to learn about the Centralized Testing program, including university courses/workshops on the content and skills measured by the examination, nonuniversity courses and workshops on examination topics, funding for attending international programs, courses and workshops that cover topics relevant to the examination, presentations, and publications (e.g., articles, collections of materials for Centralized Testing). According to the data provided in the completed SABER–Student Assessment questionnaire, opportunities are generally of high quality and are available to students studying in higher education institutions, existing full-time staff in the examination unit (RIKC), university professors, general secondary school educators, and specialists in the education management bodies. At the same time, there are no university graduate programs on student assessment that include topics relevant to the examination (e.g., test design, administration), university courses/workshops on examination topics other than the content and skills measured by the examination (e.g., test design, administration), or internships/short-term employment in the unit running the examination that would formally equip individuals to work on the Centralized Testing program.

Extensive materials are available to students to prepare for the examination. Most students have access to the official framework document explaining what is measured on the examination as well as examples of the types of questions that are on the examination, provided by either RIKC or some other source. Throughout the school year, three rounds of practice versions of the Centralized Testing are carried out, and higher education institutions provide courses to help students prepare for the Centralized Testing. According to the data provided in the completed SABER–Student Assessment questionnaire, these materials are considered to be of medium quality. In general, students do not face barriers to participating in the Centralized Testing program (there are no ethnic, racial, linguistic group identification, gender, or location—e.g., rural or hard-to-reach areas—barriers). Individuals with vision, hearing, or musculoskeletal disorders do not participate in the Centralized Testing because of a lack of appropriate accommodations or alternative assessment options. Instead, such individuals can participate in entrance examinations in

the institutions where they would like to be considered for admission, with the format of the examination being determined by the institution, taking into account the psychological development of the candidate.

Extensive quality assurance procedures are in place to ensure the technical quality of the Centralized Testing program. Specifically, the following formal quality assurance procedures are in place:

- All proctors or administrators are trained according to a protocol
- A standardized manual is available for examination administrators
- Questions/items/tasks are piloted before the official examination administration
- There are external observers and
- There are internal observers.

The examination is fully standardized at the system level in that its design, administration, scoring, and reporting are the same for all students. To monitor the overall impact of the Centralized Testing program on education quality and learning levels in the system, RIKC prepares an Annual Report, which provides information on the most recent round of examination results.

The Order of the President of the Republic of Belarus provides for the creation of a commission on the control of the preparation and implementation of entrance examinations into institutions of higher and secondary specialized education. The commission oversees the preparation and implementation of Centralized Testing every year. At the same time, focus groups or surveys of key stakeholders do not take place, and funding for validation studies or independent research on the Centralized Testing program is not available.

Sufficient documentation on the methods and procedures used during the Centralized Testing program is made available to the public and key stakeholders. Specifically, test specifications and information on the construction of questions and tasks is publicly available, while information on the scoring of questions or tasks, reliability, and scaling is prepared for the Ministry of Education but not made available to the public.

According to the data provided in the completed SABER–Student Assessment questionnaire, although the use of unauthorized materials such as prepared answers and notes, collusion among candidates via mobile phones, passing of papers, or the equivalent has taken place, the overall credibility of the examination was not compromised as a result of these individual improprieties.

After receiving their examination results, students may apply to tertiary education institutions, secondary specialized institutions, or vocational education institutions in Belarus. The examination results are not officially recognized by educational institutions or employers in other countries. Students who do not do well on the examination are not allowed to repeat the grade but may retake the examination in a future year.

Suggested policy options:

1. Ensure equitable access to the examination. For example:

- a. Improve the quality of the preparation materials available to students by reviewing existing materials, determining the areas that need to be improved, and implementing these improvements
- b. Increase the options available to students with disabilities for taking the examination, particularly students with vision or hearing disorders or musculoskeletal disabilities by, for example, adjusting the physical space (e.g., installing ramps for students using wheelchairs) and providing examination formats specifically adapted for students with these disabilities.

2. Increase opportunities for key stakeholders to learn about the examination. For example:

- a. Build a pipeline of qualified assessment specialists who could work on the examination in various capacities by
 - i. Introducing university graduate programs, courses, or workshops on assessment and testing topics relevant to the examination
 - ii. Providing opportunities to university students to do internships or apprenticeships at RIKC and gain hands-on

practical experience in carrying out examination-related activities (create confidentiality agreements with students to ensure examination security).

3. Introduce a variety of mechanisms to monitor the impact of the examination on the education system and students. For example:

- a. Form expert review groups
- b. Provide funding for independent research on the examination
- c. Conduct studies on the examination, such as predictive validity studies
- d. Conduct focus groups or surveys of key stakeholders.

National Large-Scale Assessment (NLSA) in Belarus

- Grade 6: Mathematics, world history
- Grade 7: Physics, biology
- Grades 5 and 8: Reading

Level of Development: ESTABLISHED

Belarus's National Large-Scale Assessment exercise—Republican Monitoring of Education Quality (“Republican Monitoring”)—has been administered annually since 2003. The Republican Monitoring exercise aims to assess overall levels of student achievement in relation to national education standards and syllabi requirements in particular school subjects. The assessment results are used to inform policy, pedagogy, and text book design; monitor education quality; and evaluate interventions to improve learning.

The Republican Monitoring exercise is carried out in accordance with the official Belarus Secondary School Education Development Program for 2007–16, as well as the annual decrees of the Ministry of Education “On Organization and Carrying Out Secondary School Education Monitoring in the Academic Year.” The latter offer specific guidance on funding, who and what should be assessed, and how results should be used, among other things. The Republican Monitoring exercise also has been guided by the national secondary school education monitoring programs for 2006–9 and 2009–14. The national secondary school education monitoring program for 2015–18 is being implemented.

The Republican Monitoring was first implemented in 2003, and until 2011 it covered grades 4, 9, and 11. In the 2012–13 and 2013–14 school years, it covered grades 3, 8, and 10. Between 2003 and 2014, students' knowledge and skills were assessed in the subjects of mathematics and Belarusian and Russian languages. The assessment also covered biology, chemistry, and physics in grades 9 and 11 (in 2003–11) and in grades 8 and 10 (in years 2012–14). In 2012 and 2013, knowledge and skills in social studies were assessed in grades 9 and 10.

In the 2014–15 school year, students' knowledge and skills were assessed in the following subjects and grades:

The Republican Monitoring exercise assesses students' knowledge and skills in a particular subject in relation to the content of the educational programs in these grades and subjects that students were to have acquired in the previous school year. For example, the Republican Monitoring test for mathematics in grade 6 assesses students' knowledge of mathematics content and skills that should have been covered in grade 5. Given that, starting in 2014, the Republican Monitoring exercise is carried out every year in different grades, and its objective is the assessment of the degree to which students in selected grades acquired the content of updated educational programs, comparative analysis of the results of the Republican Monitoring is not conducted.

The Republican Monitoring exercise is well aligned with the official learning standards and curriculum, as well as with classroom assessment activities and formal examinations. There are officially mandated reviews to ensure that the Republican Monitoring exercise measures what it is intended to measure, including regular internal and external reviews that take place during all or almost all Republican Monitoring rounds. Specifically, the instruments used for the Republican Monitoring exercise are subject to internal review in the Department of Education Quality Monitoring, a permanent unit at the National Institute of Education; informal external review in laboratories of the Research and Development Centre of the National Institute of Education; and formal external review in the Ministry of Education (in accordance with the decrees of the Minister of Education, all diagnostic and instructional materials for carrying out the Republican Monitoring exercise are to be coordinated in the Departments of the Ministry of Education).

Students are exposed to the content and skills measured by the assessment through regular course instruction at school; they also have access to samples of previously administered Republican Monitoring assessments. At the same time, preparatory information (e.g., official framework, approved sample questions) on the assessment is not provided to schools ahead of the test administration.

The Department of Education Quality Monitoring is responsible for the Republican Monitoring program, including the assessment design, administration, data processing, and reporting. The Ministry of Education ensures assessment quality and performs project management activities. In accordance with the Order of the Ministry of Education of the Republic of Belarus, committees that are responsible for education management in an oblast (1) select the coordinator for conducting the Republican Monitoring in institutions of general secondary education; (2) organize the participation of external observers (specialists from the rayon/city educational and methodological departments, specialists from the institutes of educational development) and the carrying out of the Republican Monitoring in regional education institutions; (3) organize technical support for the implementation of the Republican Monitoring in education institutions in the region (copying of materials, scanning of materials, carrying out a survey on the computer); and (4) provide verification and assessment of student responses in educational institutions and in the institutes of education quality. School coordinators, who are appointed by the director of an educational institution, are responsible for the implementation of assessment activities in the educational institutions. Temporary subject commissions, comprising subject teachers with the highest qualifications and teacher-methodologists (at the rayon level and in the institutes of educational development), are formed for the purposes of verification and evaluation of students' work. Specifically, specially selected secondary school subject teachers (top-ranked or supervisors) review and evaluate the question responses of students participating in the Republican Monitoring test. Experts from the Department of Education Quality Monitoring developed scoring guides to facilitate teachers' review and evaluation of students' responses. Some issues have been reported with the effectiveness of these teachers (for example, some teachers overlooked errors, which was noted by the Department of Education Quality Monitoring of the National Institute of Education staff who reviewed students' work); however, the overall quality of the assessment was not compromised as a result.

Funding for the Republican Monitoring program is provided by national and local government budgets and covers all core assessment activities. Research and development activities on the topic of assessment of

the quality of education are also financed through other government sources.

The Department of Education Quality Monitoring has most of the necessary resources to carry out its responsibilities in relation to the assessment program, although there is room for improvement in the availability of computers for technical staff, software, and servers. The department is appropriately staffed, and all staff possess the relevant qualifications for their roles.

Several opportunities are available to learn about the Republican Monitoring program, including university graduate programs, university and nonuniversity workshops, funding for participation in international courses on assessment, and opportunities to participate in online conferences and webinars. However, these opportunities are available to only a limited number of individuals.

A representative random sample of public school students in the target grades is selected to participate in the Republican Monitoring program. Students in private institutions of general secondary education comprise 0.07 percent of students in institutions of general secondary education in the Republic of Belarus and are therefore not included in the Republican Monitoring sample. Starting in 2014, students from all rural schools have an equal opportunity to be included in the sample of the Republican Monitoring. This required increasing the number of external observers by employing specialists from the rayon and city teaching offices of the institutes of educational development.

The Republican Monitoring program is fully standardized at the system level: assessment design, administration, scoring, and reporting are the same for all students in the same assessment round. Quality assurance procedures are in place to help ensure the quality of the assessment, including requiring that all proctors and administrators be trained according to a protocol, the use of a standardized manual for NLSA administrators, the numbering of all answer sheets, and the use of external observers (e.g., specialists from the rayon or city teaching offices of the institutes of educational development) as well as internal observers (e.g., NLSA staff observing at administration sites). At the same time, other desirable procedures, such as

piloting of assessment questions, training of scorers, and double processing of data, have yet to be put in place. The only inappropriate behavior that has been recorded as taking place is students copying from other students. This inappropriate behavior did not affect the credibility of the Republican Monitoring, and a decision was made in 2014 to increase the number of different versions of the assessment from two to four to address this issue.

Additionally, documentation is limited on the methods and procedures used during the Republican Monitoring exercise. Information on the evaluation and assessment of all questions is provided in the “Recommendations for Evaluation and Assessment of Students’ Work” document that is made available on the website of the management of the monitoring of the quality of education at the National Institute of Education on the day that the Republic Monitoring takes place. This information is made available to assessment coordinators in those educational institutions that took part in the monitoring. Documentation on test specifications also exists. No documentation is available on how test questions are constructed, how the test is assembled, or how scaling is done.

There is an official report for the Ministry of Education on the results of the Republican Monitoring exercise, based on which recommendations for enhancing education quality are prepared and published. The report provides information about the number of students who acquired the content of the taught (and assessed) subject at a high, sufficient, average, acceptable, and low level. Official results of the Republican Monitoring are made available at the Ministry of Education, in education-specific scientific-methodological journals, as well as on the national education portal, which publishes a summarized version with recommendations on how to improve the management of education quality. National results at the question/item level are also available through *Nastaunitskaya Gazeta*. Schools (including teachers), parents, and the local authorities have access to this information.

Limited mechanisms are in place to monitor the impact of the Republican Monitoring program on quality and learning in the education system. Focus group discussions are held with secondary school teachers during most assessment rounds; however, there is no

oversight committee or expert review groups to monitor the overall impact of the Republican Monitoring program, and no funding is available for independent research on the assessment.

Suggested policy options:

1. **Review the design of the Republican Monitoring program in light of the information needs of key decision makers and the intended purposes and uses of the assessment results. This review should cover:**
 - a. Primary purposes of the assessment in the eyes of key decision makers
 - b. Key desired uses of the assessment results, including the ability to compare scores over time and
 - c. Appropriateness of the current technical design of the assessment given these purposes and uses.
2. **To enhance the existing Republican Monitoring program, provide preparatory information about the program to schools ahead of each test administration. For example:**
 - a. A framework document explaining what is measured on the assessment
 - b. Examples of the types of questions that are on the assessment
 - c. Examples of the criteria used for scoring questions.
3. **To improve the existing Republican Monitoring program, enhance the communication of assessment results by creating guidelines for how results should be communicated to key stakeholder groups. This should include:**
 - a. Guidelines for the content and format of targeted publications for different stakeholder groups (for example, policy makers, schools, parents)
 - b. Advice on the types of results that should be communicated to different stakeholder groups (for example, question-level data for teachers).

4. **To enhance the existing Republican Monitoring program, continue to strengthen its technical quality by, for example:**
 - a. Piloting questions, items, and tasks before the official assessment administration (a decision has been made to conduct piloting)
 - b. Training scorers to ensure high interrater reliability
 - c. Documenting methods and procedures used in the assessment exercise, including information on
 - i. Construction of questions, items, and tasks
 - ii. Pilot activities and analysis of piloted questions, items, and tasks
 - iii. Test assembly
 - iv. Reliability and
 - v. Scaling.

International Large-Scale Assessment (ILSA) in Belarus

Level of Development: LATENT

Until recently, Belarus had never participated in an international large-scale assessment exercise. Now, however, the country is actively planning to take part in the 2018 Program for International Student Assessment (PISA 2018). The Republican Institute for Knowledge Control (RIKC) will be responsible for PISA 2018 activities in the country. Although PISA 2018 represents the first time that RIKC will carry out an international large-scale assessment exercise, the institute has prior experience carrying out assessments because its main function is to carry out the Centralized Testing program in Belarus. Funding support for PISA 2018–related activities in Belarus will be provided through a World Bank project loan.

In addition to RIKC staff, specialists who work at the National Institute of Education, the Academy of Post-diploma Education, institutes of educational development, and local educational authorities are prepared to work on PISA 2018, mainly because of their previous experience working on the Republican Monitoring of Education Quality program. These experts have degrees in pedagogy, social studies, and psychology, and many of them have completed training courses relevant to PISA, including a “Basic Theory and Methods of Pedagogical Measurements” course. National education quality assessment system experts regularly participate in webinars conducted by the Higher School of Economics Russian Training Centre at the Institute of Education in Moscow and receive information on education quality assessment from Eurasian Association for Educational Assessment (EAOKO) news bulletins.

Experts from the National Institute of Education also have experience conducting research in the area of educational assessment. In 2013 the institute, in association with the Russian National Training Foundation, conducted research to measure the information and communication competencies of grade 9 students in Belarus. The “Information and Communication Competence Assessment of Grade 9 Students in the Republic of Belarus” project was implemented within the framework of a small grants

program implemented by the Russian Centre for International Cooperation in Education Development (CICED).

Suggested policy options:

1. **Enabling Context:** According to best practices, the following may assist Belarus in building its enabling context for carrying out ILSA activities in the mid- to long term:
 - a. Prepare a formal policy document that authorizes ILSA activity and explains its role in supporting improved education quality and learning in Belarus, and make it available to the general public.
 - b. Develop a medium- to long-term funding plan for ILSA activities and ensure that sufficient funding is available to carry out all ILSA activities, particularly core activities essential to the technical integrity of the assessment and the utility of the results.
 - c. Conduct a needs assessment of organizational resources (e.g., computers, software, storage facilities, building security, communication tools) and human resources (e.g., specialists, translators) that will be needed to carry out ILSA activities, particularly core ILSA activities.
2. **System Alignment:** Belarus should focus on enhancing the alignment between its ILSA activity and other components of its education system. For example, the country should:
 - a. Ensure that opportunities to learn about the ILSA (including its design, content, and skills measured) are available to key stakeholder groups (including teachers) ahead of time.
 - b. Ensure that students have sufficient opportunities to be exposed to the content and skills measured by the ILSA ahead of time, including through regular instruction at school, and that they have practice responding to the types of question formats that PISA uses.
3. **Assessment Quality:** In the case of PISA and other ILSAs, many aspects of assessment quality are already monitored through standardized protocols and centralized quality assurance mechanisms. However, countries have latitude in how they

communicate and use their results. This is an area where it would be beneficial for the Belarus team to:

- a. Develop a plan for targeted communication of ILSA results to national stakeholder groups.
- b. Learn how other countries have communicated results to different stakeholders in ways that maximize understanding and use.
- c. Learn about and plan for policy- and pedagogically relevant analyses of the PISA data to inform future policy and planning in Belarus.

Appendix 1: Assessment Types and Their Key Differences

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

Appendix 2: Summary of the Development Levels for Each Assessment Type

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

Appendix 3: Methodology for Assigning Development Levels

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

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

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

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

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

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

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

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

6. Scores for certain key dimensions under “Enabling Context” (in the case of EXAM, NLSA, and ILSA) and under “System Alignment” (in the case of CLASS) were

set as ceiling scores; that is, the overall mean score for the particular assessment type cannot be greater than the score for these key dimensions. These key variables include formal policy, regular funding, having a permanent assessment unit, and the quality of assessment practices.

Appendix 4: Stakeholders Consulted during Completion of the SABER–Student Assessment Questionnaires in Belarus

Classroom Assessment Questionnaire

- Deputy Director for Educational Activities, Secondary School no. 161, Minsk
- Head of Department for Methodology of Education in the Sphere of History and Social Studies and Socio-Cultural Sphere, National Institute of Education
- Deputy Head of Education Quality Monitoring Department, National Institute of Education
- Head of Department of Auxiliary Historical Disciplines and Methods of Teaching History, Belarus State Pedagogical University n.a. Maksim Tank
- Director, Domatkanovichy Secondary School of Kletsky Rayon
- Deputy Director, Kartsevichy Secondary School of Nesvizh Rayon
- Deputy Director, Slutsk Secondary School no. 2
- Director, Krupitskaya Secondary School of Minsk Rayon
- Top-ranked methodologist in charge of professional improvement of chemistry teachers, Department for Education and Methodology, Academy of Postgraduate Education
- Mathematics teacher, Head of Subject and Methodology Division, Secondary School no. 161, Minsk
- History teacher, member of the national club, “Crystal Stork,” Secondary School no. 161, Minsk
- Deputy Director for Education, Secondary School no. 183, Minsk
- Winner of the national professional excellence competition, chairperson of the “Crystal Stork” club, Biology teacher, Smorgan Gymnasium no. 4
- Mother of 2015 graduate of Nesvizh Secondary School no. 4
- Teacher of Russian language and literature, Nesvizh Secondary School no. 4

- Chemistry teacher, leader of the oblast and rayon creative and methodological group of teachers of chemistry, Belozersk Secondary School no. 1
- Chemistry teacher and methodologist, head of rayon methodological society of teachers of science, Brest Liceum no. 1 n.a. A.S. Pushkin
- Top-rank chemistry teacher, head of methodological society of teachers of science, oligorsk Gimnazium no. 3
- Top-rank methodologist, in charge of professional improvement of teachers of mathematics, Department for Education and Methodology, Academy of Postgraduate Education
- Top-rank methodologist, in charge of professional improvement of teachers of foreign languages, Department for Education and Methodology, Academy of Postgraduate Education

Examinations Questionnaire

- Head of the General Secondary Education Department, Ministry of Education of the Republic of Belarus
- Director, Republican Institute for Knowledge Control
- First Deputy Director, Republican Institute for Knowledge Control
- Head of the Department for Monitoring of the Quality of Education, National Institute of Education
- Graduate of the 2008 PEE Secondary School no. 161, Minsk
- Graduate of the 2011 PEE Secondary School no. 161, Minsk

National Large-Scale Assessment Questionnaire

- Head of Secondary School Education Department, Ministry of Education of the Republic of Belarus
- Head of Department of Education Quality Monitoring, National Institute of Education
- Deputy Head of Department of Educational Quality Monitoring, National Institute of Education

- Leading researcher of Laboratory of Mathematics and Natural Science Education, National Institute of Education
- Head of Department of Education Sociology and Continuing Education, Academy of Postdiploma Education
- Head of Department of Cooperation with Education Sector Stakeholders, Grodno Oblast Institute of Education Development
- Head of Department of Auxiliary Sciences of History and Methods of Teaching History, Maxim Tank Belarusian State Pedagogical University
- Deputy Head of the Education Department, Moskovskiy District administration of Minsk
- Head of the Pre-school and Secondary School Department, Moskovskiy District of Minsk
- First Rank Administration Methodologist of teaching and methodology of SEU Academy of Postdiploma Education in charge of chemistry teachers' continuing education
- Deputy Director for Academic Work of SEU Minsk Grammar School no. 174

International Large-Scale Assessment Questionnaire

- Head of Secondary Education Department, Ministry of Education of the Republic of Belarus
- Director, Republican Institute for Knowledge Control
- Head of Education Quality Monitoring Department, National Institute of Education

Appendix 5: SABER–Student Assessment Rubrics for Belarus

This appendix provides the completed SABER–Student Assessment rubrics for each type of assessment activity in Belarus. In each row of the rubric, the relevant selection is indicated by a shaded cell. The selection may include a superscript number that refers to the justification or explanation for the selection. The explanation or justification text is located in the “Development-level rating justifications” section at the end of each rubric. If a row includes a superscript, but no shading, this means that insufficient information was available to determine the relevant selection in the row.

Belarus *Classroom Assessment*

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Curriculum/Standards	There was <u>no official document</u> at the system level that outlined what students were expected to learn.	There was an <u>official document</u> at the system level, but it provided <u>limited and insufficient information</u> on what students were expected to learn.	There was an <u>official document</u> at the system level that provided <u>sufficient, but not extensive information</u> on what students were expected to learn.	There was an <u>official document</u> at the system level that provided <u>extensive and comprehensive information</u> on what students were expected to learn. ¹
Policy Document	There was <u>no document</u> at the system level that provided <u>guidelines</u> for classroom assessment.	There was a <u>document</u> at the system level that provided <u>guidelines</u> for classroom assessment, but it was either <u>unofficial, not publicly available, or provided limited guidance.</u>	There was an <u>official and publicly available document</u> at the system level that provided <u>sufficient, but not extensive, guidelines</u> for classroom assessment.	There was an <u>official and publicly available document</u> at the system level that provided <u>extensive and comprehensive guidelines</u> for classroom assessment. ²
Resources	There were <u>no resources</u> available to teachers in the system for their use in classroom assessment activities.	There were <u>resources available</u> to teachers in the system for their use in classroom assessment activities, but these resources were <u>not of high quality and were limited</u> in number or availability.	There were <u>sufficient high-quality resources</u> available to all or almost all teachers in the system for their use in classroom assessment activities.	There was an <u>extensive number of high-quality resources</u> available to all or almost all teachers in the system for their use in classroom assessment activities. ³

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Teacher Development	There were <u>no formal mechanisms</u> at the system level that supported the development of teachers' competencies in classroom assessment.	There was a <u>minimum number of formal mechanisms</u> at the system level that supported the development of teachers' competencies in classroom assessment, or else the available formal mechanisms were <u>not of high quality or were limited</u> in their availability.	There were <u>sufficient formal mechanisms of high quality</u> at the system level that supported the development of teachers' competencies in classroom assessment. ⁴	There was an <u>extensive number of high-quality, formal mechanisms</u> at the system level that supported the development of teachers' competencies in classroom assessment.
Quality Monitoring	There were <u>no formal mechanisms</u> at the system level to monitor the quality of classroom assessment practices.	There was a <u>minimum number of formal mechanisms</u> at the system level to monitor the quality of classroom assessment practices.	There were <u>sufficient formal mechanisms</u> at the system level to monitor the quality of classroom assessment practices, including inspection/supervision. ⁵	There were <u>extensive formal mechanisms</u> at the system level to monitor the quality of classroom assessment practices, including inspection/supervision.
Report to Stakeholders	Schools were <u>not required</u> to report individual student performance on classroom assessments to any stakeholders.	Schools had <u>minimum requirements to report</u> individual student performance on classroom assessments.	Schools were <u>required to report</u> individual student performance on classroom assessments to the student and their parents. ⁶	Schools were <u>required to report</u> individual student performance on classroom assessments to a variety of relevant stakeholders.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Report Content	Schools were <u>not required</u> to report <u>individual</u> student performance on classroom assessments in particular subject areas.	Schools were required to report individual student performance on classroom assessments in <u>one or two subject areas</u> .	Schools were required to report individual student performance on classroom assessments in <u>more than two subject areas</u> . ⁷	This option does not apply to this indicator.
Report Format	There were <u>no requirements</u> for schools to use specific formats for reporting individual student performance on classroom assessments.	There were <u>requirements</u> for schools to use specific formats for reporting individual student performance on classroom assessments to the students and their parents, but the formats specified <u>did not include written reports and teacher/school meetings</u> . ⁸	There were requirements for schools to use specific formats for reporting individual student performance on classroom assessments to the students and their parents, including <u>written reports and teacher/school meetings</u> . There also were requirements for schools to report this information to the school district, Ministry of Education, or equivalent although the format for reporting to these entities was not specified.	There were requirements for schools to use specific formats for reporting individual student performance on classroom assessments to students, parents, and the school district, Ministry of Education, or equivalent. These included <u>written reports and (in the case of students and parents) teacher/school meetings</u> .
Required Uses	There were <u>no system-level</u> requirements for teachers to use classroom assessment information.	Teachers were required to use classroom assessment information, albeit in a <u>minimal number of ways</u> .	Teachers were required to use classroom assessment information in a <u>sufficient number of ways</u> .	Teachers were required to use classroom assessment information in <u>extensive ways</u> . ⁹

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Selection and Certification	At the secondary level, classroom assessment information was <u>not</u> required as an <u>input</u> for certification decisions or for selection to the next level of the education system.	This option does not apply to this indicator.	At the secondary level, classroom assessment information was <u>required to be used as an input</u> for certification decisions or for selection to the next level of the education system. ¹⁰	This option does not apply to this indicator.
Positive Uses	Classroom assessment information was <u>used in positive ways</u> by a <u>marginal number</u> of teachers.	Classroom assessment information was <u>used in positive ways</u> by <u>some</u> teachers.	Classroom assessment information was <u>used in positive ways</u> by <u>most</u> teachers. ¹¹	Classroom assessment information was <u>used in positive ways</u> by <u>all or almost all</u> teachers.
Poor Practices	<u>All or almost all</u> teachers engaged in <u>poor</u> classroom assessment practices.	<u>Many</u> teachers engaged in <u>poor</u> classroom assessment practices.	<u>Only some</u> teachers engaged in <u>poor</u> classroom assessment practices.	<u>A marginal number</u> of teachers or <u>no</u> teachers engaged in <u>poor</u> classroom assessment practices. ¹²

Classroom Assessment: Development-level rating justifications

1. “Educational Standards for School Subjects, 2009” is an official learning standards document at the system level that outlines what students at different grade or age levels are expected to learn, but does not specify the desired performance level. The document covers subjects such as language (Belarusian and Russian), mathematics, biology, geography, world history, history of Belarus, Russian literature, Belarusian literature. Official system-level documents also include “Educational Standards for School Subjects” (2012, 2015), which outlines what students at different grade or age levels are expected to learn and the desired performance level, and the “Educational Programs for Extracurricular Subjects, 2009–2013,” which outlines what students at different grade or age levels are expected to learn but does not specify the desired performance level.
2. Formal, official documents at the system level that provide guidance for classroom assessment and are available to the general public include:
 - “Assessment Norms for Outcomes of Educational Activity of Students in Educational Subjects”: this document describes (a) what should be assessed, (b) criteria for scoring student work, (c) formats and methods for carrying out classroom assessment activities, (d) description of significant and insignificant errors in student responses, (e) performance evaluation of students on a 10-point scale, (f) assessment tests and tasks scale to translate points for the test into marks, (g) mechanisms for issuing test scores for subjects.
 - Instructions and methodological letter “On Organization of Work of General Secondary Education Institutions to Monitor and Evaluate the Outcomes of Learning Activity of Students during the Ungraded Learning Stage I of General Secondary Education”: this document describes (a) what should be assessed, (b) principles of ungraded training, (c) a mechanism to evaluate the learning process, (d) types and forms of control under ungraded training, (e) methods of formation of adequate student self-esteem, (f) record keeping by teachers and students, and administration of educational institutions, (g) rights and responsibilities of students and teachers, and the legal representatives of pupils.
 - “Rules for Certification of Pupils at Mastering Content Educational Programs of General Secondary Education”: this document describes (a) uses of assessment information, (b) the form of certification in secondary education institutions, (c) procedure for promotion to the next grade, (d) procedure for revising the annual grade.
3. Extensive resources are available to teachers in the system for their classroom assessment activities.

Available resources include:

 - A. A document that outlines what students are expected to learn in different assessment domains at different grade/age levels
 - B. A document that outlines the performance level(s) that students are expected to reach in different assessment domains at different grade/age levels
 - C. Student textbooks that provide support for classroom assessment (all teachers receive textbooks that contain questions and sample material for conducting classroom assessment activities).
 - D. Teacher guides (compilations of materials to be used for conducting continuous, intermediate, and final assessment have been published for all school subjects in grades 1–11 and are available for purchase; all subject teachers receive funds specifically to acquire the methodological literature).
 - E. Scoring criteria or rubrics for grading student work (diagnostic assessment modules for all subjects in grades 1–11 are available on the national education portal). All participants in the education process, including students, parents, and teachers, have access to these resources.

- F. Item banks/pools with examples of questions/tasks to be used for classroom assessment activities
- G. Computer-based classroom assessment resources (diagnostic assessment modules have been developed for all subjects in grades 1–11 and made available on the national education portal. All stakeholders in the education process, including students, parents, and teachers, have free access to these modules).

Resource A, B, and E are of high quality and available to all or almost all teachers

Resource C is of medium quality and available to all or almost all teachers

Resource D is of high quality and available to most teachers

Resource F is of high quality and available to all or almost all teachers

Resource G is of high quality and available to all or almost all teachers.

4. Extensive high-quality mechanisms are in place at the system level to support the development of teachers' competencies in classroom assessment and are available on an annual basis.

Available mechanisms include:

- A. Preservice teacher training that addresses competencies in classroom assessment. This mechanism is available to all or almost all teachers, although not on an annual basis. This mechanism is considered to be of medium quality.
- B. In-service teacher training that addresses competencies in classroom assessment. This mechanism is available to all or almost all teachers, though not on an annual basis, and is considered to be of high quality.
- C. Online resources on classroom assessment. This mechanism is available to all or almost all teachers on an annual basis and is considered to be of high quality.
- D. Opportunities to participate in conferences and workshops on classroom assessment. This mechanism is available to all or almost all teachers on an annual basis and is considered to be of medium quality.
- E. Opportunities to participate in the development or scoring of test questions for large-scale assessments. This mechanism is considered to be of high quality but is available to only a marginal number of teachers and not on an annual basis.
- F. Implementation of experimental and innovative projects on piloting and introduction of new technologies for evaluation of students' achievements. This mechanism is considered to be of high quality but is available to only a marginal number of teachers and not on an annual basis.
- G. Distance learning for teachers in classroom assessment. This mechanism is considered to be of high quality and is available to only a marginal number of teachers on an annual basis.
- H. Self-education of teachers in classroom assessment. This mechanism is available to all teachers on an annual basis but is considered to be of medium quality.

Although all or almost all (more than 90 percent) teachers can participate in A, B, C, D, and H, a marginal number (less than 10 percent) of teachers can participate in E, F, and G.

5. Formal mechanisms exist at the system level to monitor the quality of classroom assessment practices. Four mechanisms are found, which is less than extensive (five or six mechanisms) and more than minimal (one or two mechanisms).
Mechanisms include the following: (a) Classroom assessment is a required component of school inspection/teacher supervision, (b) classroom assessment is a required component of a teacher's performance evaluation, (d) regular, high-level reviews of the quality of education include a focus on classroom assessment, and (e) government funding is awarded to researchers and practitioners for research and development on classroom assessment practices. An external moderation system that consists of commissions in educational subjects at the district level that objectively evaluate the assessment of written examination questions of those students who are being considered to receive an attestation of general secondary education with an award of a gold (or silver) medal is also in place. At the same time, a more wide-reaching moderation system, or a moderation system where teachers from different schools meet to review assessment tasks, scoring criteria, and students' work, with the purpose of ensuring the validity and comparability of student scores and grades is not in place, is not in place.
6. All or almost all schools report individual student performance on classroom assessment in all subject areas to students and their parents through school-wide and class-specific parent-teacher conferences. Student grades are recorded in a classroom journal and in the student's grade journal (*dnevnik*). Parents are required to review the student's grade journal every week and every quarter. There is currently a move toward maintaining the classroom journal and the student's grade journal in electronic form. Schools are not required to report to school district, Ministry of Education, or equivalent.
7. Schools are required to report individual student performance in all subjects, including language and mathematics.
8. Schools report individual student performance on classroom assessment in all subject areas to students and their parents through schoolwide and class-specific parent-teacher conferences. Student grades are recorded in a classroom journal and in the student's grade journal (*dnevnik*). Parents are required to review the student's grade journal every week and every quarter. There is currently a move toward maintaining the classroom journal and the student's grade journal in electronic form. Schools are not required to report individual student performance information to school district, Ministry of Education, or equivalent entity in any format, including written and electronic.
9. Classroom assessment is required to be used in an extensive number (four) of ways by teachers, specifically, to:
 - Diagnose student learning issues
 - Provide continuous feedback to students as part of instruction
 - Plan the next steps in instruction
 - Evaluate student performance
10. Classroom assessment information is required to be used for certification and as an input for selection to the next level within the education system.

11. Classroom assessment information is used in positive ways by most (50–90 percent) teachers. Specifically, teachers use classroom assessment information to:

- Diagnose student learning issues (most teachers)
- Provide continuous feedback to students as part of instruction (most teachers)
- Plan the next steps in instruction (most teachers)
- Evaluate student performance (all or almost all teachers)

12. Overall, a marginal number of teachers engage in poor classroom assessment practices.

No teachers use assessment tools that are not aligned with the pedagogical/curriculum framework; a marginal number of teachers (less than 10 percent) apply uneven standards for grading student work; some teachers inflate grades; some teachers overemphasize memorization/recall of information; a marginal number of teachers (less than 10 percent) overemphasize the use of multiple choice/selection-type questions; an insignificant number of teachers (less than 5 percent) make errors in the scoring or grading of student work.

BELARUS *Examinations*

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Program Stability	No <u>examination program</u> existed at the system level.	An <u>examination program</u> existed at the system level, but it was <u>not sufficiently stable</u> .	A <u>stable examination program</u> had been in place for <u>several years</u> .	A <u>stable examination program</u> had been in place for <u>10 years</u> or more. ¹
Clarity of Purpose	There were <u>no policy-mandated purposes</u> of the examination.	The examination had clear <u>policy-mandated purposes</u> , but these <u>did not include</u> student certification or selection.	The examination had clear <u>policy-mandated purposes</u> that <u>included</u> student certification, selection, or both. ²	This option does not apply to this indicator.
Policy Document	No <u>policy document</u> authorized the examination program.	An <u>informal/draft policy document</u> authorized the examination program.	A <u>formal/official policy document</u> authorized the examination program, but the document was <u>not available</u> to the general public.	A <u>formal/official policy document</u> authorized the examination program and was <u>available</u> to the general public. ³
Program Guidelines	No <u>official document</u> provided guidelines for the examination program.	An <u>official document</u> provided guidelines for the examination program, but it was <u>missing some</u> key guidelines.	An <u>official document</u> provided <u>key</u> guidelines for the examination program. ⁴	This option does not apply to this indicator.
Stability of Organization	There was <u>no unit</u> with primary responsibility for running the examination program.	There was a <u>unit(s)</u> with primary responsibility for running the examination program, but the <u>unit(s)</u> was <u>temporary</u> or had been <u>in place for less than 5 years</u> .	There was a <u>permanent unit(s)</u> with primary responsibility for running the examination program that <u>had</u> been in place for 5 or more years. ⁵	This option does not apply to this indicator.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Accountability of Organization	There was <u>no</u> unit with primary responsibility for running the examination program, or else the unit responsible was <u>not</u> accountable to a clearly recognized body.	The <u>unit(s)</u> with primary responsibility for running the examination program was <u>accountable to a clearly recognized body within the examination unit.</u>	The <u>unit(s)</u> with primary responsibility for running the examination program was <u>accountable to a clearly recognized body within the same institution as the examination unit.</u>	The <u>unit(s)</u> with primary responsibility for running the examination program was <u>accountable to a clearly recognized external body.</u> ⁶
Organization Resources	The examination unit <u>did not have</u> the appropriate resources.	The examination unit <u>had some of</u> the appropriate resources.	The examination unit <u>had most of</u> the appropriate resources.	The examination unit <u>had all of</u> the appropriate resources. ⁷
Qualifications of Staff	There were <u>no</u> individuals responsible for completing key examination activities.	<u>Some</u> of the individuals responsible for completing key examination activities <u>had the relevant qualifications.</u>	<u>Most</u> of the individuals responsible for completing key examination activities <u>had the relevant qualifications.</u>	<u>All or almost all</u> of the individuals responsible for completing key examination activities <u>had the relevant qualifications.</u> ⁸
Effectiveness of Staff	There were <u>no</u> individuals responsible for completing key examination activities.	The responsible individuals completed key examination activities, but there were <u>significant issues</u> in how these activities were completed.	The responsible individuals completed key examination activities, with <u>only some issues</u> in how these activities were completed.	The responsible individuals completed key examination activities, and there were <u>no issues</u> in how these activities were completed. ⁹
Source of Funding	There was <u>no</u> funding available for examination activities.	The source of <u>funding</u> for the majority of examination activities was <u>loans, credits, grants, or equivalent.</u>	The source of <u>funding</u> for the majority of examination activities was the <u>government's internal funding sources or student fees.</u> ¹⁰	This option does not apply to this indicator.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Activities Funded	There was <u>no funding</u> available for examination activities.	Funding was <u>not sufficient</u> to cover all core examination activities.	Funding was <u>sufficient</u> to cover all core examination activities. ¹¹	This option does not apply to this indicator.
Staff/Teacher Opportunity to Learn	There were <u>no opportunities</u> to learn about the examination.	<u>Opportunities</u> to learn about the examination were <u>minimal, or not of high quality, or did not benefit all key stakeholder groups</u> . ¹²	There were <u>sufficient high-quality opportunities</u> to learn about the examination that were available to <u>key stakeholder groups</u> .	<u>Opportunities</u> to learn about the examination were <u>extensive, of high quality, and benefited key stakeholder groups</u> .
Teacher Participation	Teachers <u>did not perform</u> examination-related tasks. ¹³	Teachers <u>performed a minimal number</u> of examination-related tasks.	Teachers <u>performed a sufficient number</u> of examination-related tasks.	Teachers <u>performed an extensive number</u> of examination-related tasks.
Measuring What Intended	It was <u>not clear</u> what the examination was intended to measure.	There was <u>weak alignment</u> between the examination and what it was meant to measure, or there was <u>no regular review</u> process in place to verify that alignment existed.	The examination measured <u>official learning standards or curriculum</u> , and <u>officially mandated reviews</u> to verify this alignment took place during <u>most examination rounds</u> . ¹⁴	The examination measured <u>official learning standards or curriculum</u> , and <u>officially mandated reviews</u> to verify this alignment took place during <u>all or almost all examination rounds</u> . ¹⁴
Alignment with Other Assessments	The examination was <u>poorly aligned</u> with other types of assessment activities in the system.	The examination was somewhat <u>aligned</u> with other types of assessment activities in the system.	The examination was <u>very aligned</u> with other types of assessment activities in the system. ¹⁵	This option does not apply to this indicator.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Availability of Preparation Materials	There were <u>no materials</u> available to students to prepare for the examination.	Materials to prepare for the examination were <u>available to some or a marginal number</u> of students.	Materials to prepare for the examination were <u>available to most</u> students. ¹⁶	Materials to prepare for the examination were <u>available to all or almost all</u> students.
Quality of Preparation Materials	There were <u>no materials</u> available to students to prepare for the examination.	<u>Minimal material</u> was available to students to prepare for the examination, or the material available was <u>not of high quality</u> . ¹⁷	<u>Sufficient and high-quality material</u> was available to students to prepare for the examination.	<u>Extensive and high-quality material</u> was available to students to prepare for the examination.
Reasons for Not Taking the Examination	All or <u>almost all individuals</u> could not take the examination due to <u>one or more non-examination-relevant reason(s)</u> .	<u>Most or some individuals</u> could not take the examination due to <u>one or more non-examination-relevant reason(s)</u> .	There were <u>no non-examination-relevant reasons</u> that prevented individuals from taking the examination. ¹⁸	This option does not apply to this indicator.
Quality Assurance	<u>No formal procedures</u> were in place to ensure the quality of the examination.	<u>Formal procedures</u> to ensure the quality of the examination were <u>minimal in nature or not required</u> .	<u>Formal procedures</u> to ensure the quality of the examination were <u>sufficient in nature and required</u> .	<u>Formal procedures</u> to ensure the quality of the examination were <u>extensive in nature and required</u> . ¹⁹
Standardization	The examination was <u>not</u> standardized at the system level.	The examination was <u>partially</u> standardized at the system level, or <u>minimal or no procedures</u> were in place to ensure standardization.	The examination was <u>fully</u> standardized at the system level, and <u>sufficient procedures</u> were in place to ensure standardization.	The examination was <u>fully</u> standardized at the system level, and <u>extensive procedures</u> were in place to ensure standardization. ²⁰

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Quality Processes	<u>Many errors or delays</u> in activities took place that affected the examination <u>to a great extent</u> .	<u>Errors or delays</u> in activities affected the examination <u>to a significant level</u> .	Any <u>errors or delays</u> in activities had <u>only a minimal effect</u> on the examination.	<u>Errors or delays</u> in activities <u>did not affect</u> the examination. ²¹
Inappropriate Behavior	Inappropriate behavior <u>compromised the credibility</u> of the examination <u>to a great extent</u> .	Inappropriate behavior took place and <u>compromised the credibility</u> of the examination <u>somewhat</u> .	Inappropriate behavior was <u>low</u> and <u>did not compromise the credibility</u> of the examination.	Inappropriate behavior, if any, was <u>marginal and did not compromise the credibility</u> of the examination. ²²
Credibility of Results	The results of the examination were perceived as <u>credible</u> by <u>very few stakeholder groups</u> .	The results of the examination were perceived as <u>credible</u> by <u>some stakeholder groups</u> .	The results of the examination were perceived as <u>credible</u> by <u>most stakeholder groups</u> .	The results of the examination were <u>perceived as credible</u> by <u>all or almost all stakeholder groups</u> . ²³
Confidentiality of Results	There was <u>no official policy</u> to keep student results <u>confidential</u> , and student results were <u>not kept confidential in practice</u> .	Confidentiality of student results was <u>partially accomplished</u> .	There was an <u>official policy</u> to keep student results <u>confidential</u> , and student results were <u>kept confidential in practice</u> . ²⁴	This option does not apply to this indicator.
Official Recognition of Results	Examination results were <u>not officially recognized</u> by educational institutions or employers in other countries. ²⁵	This option does not apply to this indicator.	Examination results were <u>officially recognized</u> by educational institutions or employers in other countries.	This option does not apply to this indicator.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Post-Examination Options for Students	No <u>options</u> were available to students after they had taken the examination.	<u>Minimal options</u> were available to students after they had taken the examination.	<u>Sufficient options</u> were available to students after they had taken the examination. ²⁶	<u>Extensive options</u> were available to students after they had taken the examination.
Methods and Procedures Documentation	There was <u>no documentation</u> on the methods and procedures used during the examination.	There was <u>minimal documentation</u> on the methods and procedures used during the examination, or the documentation that existed was not public.	There was <u>sufficient and public documentation</u> on the methods and procedures used during the examination. ²⁷	There was <u>extensive and public documentation</u> on the methods and procedures used during the examination.
Impact Monitoring	No <u>mechanisms</u> were in place to monitor the impact of the examination.	<u>Minimal mechanisms</u> were in place to monitor the consequences of the examination, or the mechanisms took place in <u>only some or a few examination rounds</u> . ²⁸	<u>Sufficient mechanisms</u> were in place to monitor the impact of the examination and the mechanisms took place <u>all or almost all examination rounds</u> .	<u>Extensive mechanisms</u> were in place to monitor the impact of the examination and the mechanisms took place <u>all or almost all examination rounds</u> .
Readiness to Start Examination Program	The system was <u>weakly prepared</u> to start an examination program in the future.	The system was <u>somewhat prepared</u> to start an examination program in the future.	The system was <u>well prepared</u> to start an examination program in the future.	This option does not apply to this indicator. ²⁹

Examinations: Development-level rating justifications

1. “Centralized Testing” is an examination program that exists at the system level. This examination was administered for the first time more than 10 years ago (in 2003) and has since been administered annually.
2. Results from Belarus’s national examination program, “Centralized Testing,” are used to select individuals into higher and specialized secondary education institutions. Centralized Testing can be taken by those individuals who have completed the educational program of general secondary education, of the professional-technical education that provides vocational education and general secondary education, or of specialized secondary education as confirmed by a document of completion of such education. Centralized Testing is also used to inform policy and for evaluating interventions aimed at improving student learning.
3. The Resolution of the Council of Ministers of the Republic of Belarus no. 714, “On Approval of the Regulation on the Procedure for Organizing and Holding Centralized Testing” (June 6, 2006), is the formal, publicly available, system-level policy document that authorizes the examination program.
4. A number of formal, publicly available system-level documents provide guidelines on the examination, including what should be assessed, who should be assessed, and how the results should be communicated to stakeholders. These documents include the Resolution of the Council of Ministers of the Republic of Belarus no. 714, “On Approval of the Regulation on the Procedure for Organizing and Holding Centralized Testing” (June 6, 2006); “Rules of Admission of Persons for the 1st Stage of Higher Education (as amended by Presidential Decree of 03.20.2014, no. 130)” ; the Resolution of the Council of Ministers of the Republic of Belarus no. 19, “On Conducting the Centralized Testing in Academic Subjects in 2015” (April 13, 2015); and the Order of the Minister of Education of the Republic of Belarus no. 323, “On Organization of Centralized Testing in 2015” (April 22, 2015).
5. The unit with primary responsibility for running the examination program is the Republican Institute for Knowledge Control (RIKC), a permanent agency that took charge of the examination more than 10 years ago. Specifically, since 2000, RIKC has been responsible for examination design, administration, project management, data processing, and research and development. RIKC is accountable to the Ministry of Education.
6. The Ministry of Education also provides guidance on the rules of computing examination scores and the organizational rules at the centers carrying out the Centralized Testing. The Ministry of Education is subordinate to the Council of Ministers of the Republic of Belarus.
7. RIKC, which has the primary responsibility for running the Centralized Testing program, is accountable to the Ministry of Education, from which it is institutionally separate.
7. RIKC currently has all of the appropriate resources, including appropriate computers for all technical staff, software, building security, storage facilities, computer servers, and communication tools (phone, email, Internet).

8. All or almost all (more than 90 percent) of the individuals currently responsible for completing key examination activities have relevant qualifications.

Individuals responsible for completing key examination activities include:

- Permanent staff from the examination unit
- Temporary staff hired to support a particular examination cycle (e.g., examination administrators)

9. No issues are found with the effectiveness of any of the staff, or else minor issues were identified that have had no consequences for the quality of specific examination activities or for the overall quality of the examination.

10. Funding is available for examination activities. The source of funding for the majority of examination activities is government funding.

Funding for examination activities supports examination design, administration, data processing, data reporting, project management, and research and development activities. These activities are supported by the government budget. Data processing is further supported by RIKC extra-budgetary funds.

11. Funding is sufficient to cover all core examination activities as well as research and development.

12. There are sufficient opportunities to learn about the examination. Opportunities to learn did not include university graduate programs.

Opportunities to learn about examinations include university courses/workshops on the content and skills measured by the examination; non-university courses/workshops on examination topics; funding for attending international programs/courses/workshops on student assessment that cover topics relevant to the examination; presentations; instruction and methodology meetings with members of organizing committees working in venues of centralized testing; publications (articles, collections of materials for centralized testing).

According to the data provided in the completed SABER–Student Assessment questionnaire, opportunities are generally of high quality and are available to students studying in higher education institutions, existing full-time staff in the examination unit (RIKC), university professors, general secondary school educators, and specialists in the education management bodies.

The following opportunities were not available in the system to learn about the examination: university graduate programs (master’s or doctorate level) on student assessment that include topics relevant to the examination (e.g., test design, administration); non-university courses/workshops on the content and skills measured by the examination (e.g., courses on curriculum); university courses/workshops on examination topics other than the content and skills measured by the examination (e.g., test design, administration); internships/short-term employment in the unit running the examination.

13. Teachers do not perform examination-related tasks.

14. The Centralized Testing program is intended to measure official learning standards and curriculum. In practice, the examination is very aligned with what it is intended to measure. Officially mandated reviews to ensure that the examination measures what is intended to measure take place in the form of regular external reviews and regular internal reviews, both of which take place in all or almost all examination rounds (more than 90 percent).
15. The Centralized Testing program is closely aligned with Classroom Assessments and National Large-Scale Assessments in Belarus in terms of the content being measured.
16. Materials were available to students to prepare for the examination. On average, these materials were available to most (50 to 90 percent) students.
- Available materials included (a) the official framework document explaining what was measured on the examination; (b) examples of the types of questions that were on the examination, provided by RIKC; (c) examples of the types of questions that were on the examination, provided by an entity other than RIKC; (d) information on how to prepare for the examination, provided by RIKC; and (e) information on how to prepare for the examination, provided by an entity other than RIKC, such as by higher education institutions.
- A, b, and d were available to all or almost all students; c and e were available to some students.
17. Although the number of materials available to prepare for the examination was extensive, material content was of “medium” quality overall.
18. Individuals with the following profiles do not face barriers that might prevent otherwise eligible individuals from taking the examination: ethnic, racial, or linguistic group identification; gender; or location (e.g., rural or hard-to-reach areas). Individuals with vision, hearing, or musculoskeletal disorders do not take part in the Centralized Testing program. Instead, such individuals can participate in entrance examinations in the institutions where they would like to be considered for admission, with the format of the examination being determined by the institution taking into account the psychological development of the candidate.
19. Formal quality assurance procedures were in place to ensure the quality of the examination. These procedures were extensive in number (six or seven), and all procedures were required to take place every examination round.
- The following formal quality assurance procedures were in place to ensure the quality of the examination:
- All proctors or administrators were trained according to a protocol
 - There was a standardized manual for examination administrators
 - Questions/items/tasks were piloted before the official examination administration
 - There were external observers
 - There were internal observers

- A process was outlined by which control of the organization and conduct of Centralized Testing is performed by the State Control Commission for the preparation and conduct of entrance examinations to institutions of higher and secondary specialized education, as approved by a Presidential Edict.

No double processing of data was in place.

20. The examination was fully standardized at the system level. Assessment design, administration, scoring, and reporting were the same (or equivalent) for all students in the examination round.

Extensive procedures (five or six) were in place to ensure the standardization of the examination at the system level including:

- Examination papers and tasks were the same or equivalent for all students
- Examination administrators were trained to ensure that all students took the examination under the same conditions
- Quality control monitors/observers were used to ensure the same administration conditions in all locations where the examination was administered
- The same scoring criteria were used to correct the examination questions/items/tasks
- Examination results were computed using the same procedures for all students
- Examination results were reported to all students in the same way

21. No issues affected the examination round. There were instances of errors in the printing of test booklets; however, this did not affect the examination round.

22. Although the use of unauthorized materials such as prepared answers and notes as well as collusion among candidates via mobile phones, passing of papers, or equivalent did take place, the credibility of the examination was not compromised at all because of these inappropriate behaviors.

23. All or almost all stakeholder groups perceive the examination results as credible.

24. Students' Centralized Testing results were kept confidential, in line with official policy.

25. Examination results were not officially recognized by educational institutions or employers in other countries.

26. Sufficient options are available to students after they take the examination and receive their results. Students can apply to tertiary education institutions, secondary specialized institutions, or vocational education institutions. Individuals may also retake the examination in a future year.

27. There was documentation on the methods and procedures used during the examination. This documentation was sufficient and for the most part available to the general public.

Documented procedures include:

- Test specifications (publicly available)
- Construction of questions/items/tasks (publicly available)
- Test assembly (publicly available)
- Scoring of examination questions/items/tasks (publicly available)
- Reliability (not publicly available)
- Scaling (not publicly available)

The minimum scores for the Centralized Testing (which allow for individuals to apply to intuitions of higher education) are provided on an annual basis by the Ministry of Education. The following aspects of the examination are not documented: pilot testing of questions/items/tasks; analysis of piloted questions/items/tasks; marking/scoring of open-ended questions/items/tasks.

28. To monitor the impact of the Centralized Testing, RIKC produces an Annual Report for all examination rounds. Additionally, the Order of the President of the Republic of Belarus provides for the creation of a commission on the control of the preparation and implementation of entrance examinations into institutions of higher and secondary specialized education. The commission oversees the preparation and implementation of Centralized Testing every year.
29. This indicator does not apply to this rubric.

BELARUS

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

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Program Stability	No <u>NLSA program</u> existed at the system level.	An <u>NLSA program</u> existed at the system level, but it was <u>not sufficiently stable</u> .	A <u>stable NLSA program</u> had been in place for <u>several years</u> .	A <u>stable NLSA program</u> had been in place for <u>10 years or more</u> . ¹
Clarity of Purpose	There were <u>no policy-mandated purposes</u> of the NLSA.	The NLSA had clear <u>policy-mandated purposes</u> , but these <u>did not include informing policy or pedagogy</u> .	The NLSA had clear <u>policy-mandated purposes that included informing policy or pedagogy</u> . ²	This option does not apply to this indicator.
Policy Document	<u>No policy document</u> authorized the NLSA program.	An <u>informal/draft policy document</u> authorized the NLSA program.	A <u>formal/official policy document</u> authorized the NLSA program, but the document was <u>not available</u> to the general public.	A <u>formal/official policy document</u> authorized the NLSA program and was <u>available</u> to the general public. ³
Program Guidelines	<u>No official document</u> provided guidelines for the NLSA program.	An <u>official document</u> provided guidelines for the NLSA program, but it was <u>missing some key guidelines</u> .	An <u>official document</u> provided <u>key guidelines</u> for the NLSA program. ⁴	This option does not apply to this indicator.
Stability of Organization	There was <u>no unit</u> with primary responsibility for running the NLSA program.	There was a <u>unit(s)</u> with primary responsibility for running the NLSA program, but the unit(s) was <u>temporary or had been in place for less than 5 years</u> .	There was a <u>permanent unit(s)</u> with primary responsibility for running the NLSA program that had been in place for <u>5 or more years</u> . ⁵	This option does not apply to this indicator.

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Accountability of Organization	There was <u>no unit</u> with primary responsibility for running the NLSA program, or else the unit responsible was <u>not accountable</u> to a clearly recognized body.	The <u>unit(s)</u> with primary responsibility for running the NLSA program was <u>accountable</u> to a clearly recognized body <u>within the NLSA unit</u> .	The <u>unit(s)</u> with primary responsibility for running the NLSA program was <u>accountable</u> to a clearly recognized body <u>within the same institution as the NLSA unit</u> . ⁶	The <u>unit(s)</u> with primary responsibility for running the NLSA program was <u>accountable</u> to a clearly recognized external body.
Source of Funding	There was <u>no funding</u> available for NLSA activities.	The source of <u>funding</u> for the majority of NLSA activities was <u>loans, credits, grants, or equivalent</u> .	The source of <u>funding</u> for the majority of NLSA activities was the <u>government's internal funding sources</u> . ⁷	This option does not apply to this indicator.
Activities Funded	There was <u>no funding</u> available for NLSA activities.	Funding was <u>not sufficient</u> to cover all core NLSA activities.	Funding was <u>sufficient</u> to cover all core NLSA activities. ⁸	This option does not apply to this indicator.
Organization Resources	The NLSA unit <u>did not have</u> the appropriate resources.	The NLSA unit <u>had some</u> of the appropriate resources.	The NLSA unit <u>had most</u> of the appropriate resources. ⁹	The NLSA unit <u>had all</u> of the appropriate resources.

<p>Qualifications of Staff</p>	<p>There were <u>no</u> individuals responsible for completing key NLSA activities.</p>	<p><u>Some</u> of the individuals responsible for completing key NLSA activities had the <u>relevant</u> qualifications.</p>	<p><u>Most</u> of the individuals responsible for completing key NLSA activities had the <u>relevant</u> qualifications.</p>	<p>All or almost all of the individuals responsible for completing key NLSA activities had the <u>relevant</u> qualifications.¹⁰</p>
---------------------------------------	---	--	--	---

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Effectiveness of Staff	There were <u>no</u> individuals responsible for completing key NLSA activities.	The <u>responsible</u> individuals completed key NLSA activities, but there were <u>significant issues</u> in how these activities were completed.	The <u>responsible</u> individuals completed key NLSA activities, with <u>only some issues</u> in how these activities were completed. ¹¹	The <u>responsible</u> individuals completed key NLSA activities, and there were <u>no issues</u> in how these activities were completed. ¹¹
Staff/Teacher Opportunity to Learn	There were <u>no opportunities</u> to learn about the NLSA.	<u>Opportunities</u> to learn about the NLSA were <u>minimal</u> , or <u>not of high quality</u> , or <u>did not benefit all</u> key stakeholder groups.	There were <u>sufficient high-quality opportunities</u> to learn about the NLSA that were available to <u>key stakeholder groups</u> . ¹²	<u>Opportunities</u> to learn about the NLSA were <u>extensive</u> , of <u>high quality</u> , and <u>benefited key stakeholder groups</u> .
Measuring What Is Intended	It was <u>not clear</u> what the NLSA was intended to measure.	There was <u>weak alignment</u> between the NLSA and what it was meant to measure, or there was <u>no regular review process</u> in place to verify that alignment existed.	The NLSA measured <u>official learning standards or curriculum</u> , and <u>officially mandated reviews</u> to verify this alignment took place during <u>most</u> NLSA rounds.	The NLSA measured <u>official learning standards or curriculum</u> and <u>officially mandated reviews</u> to verify this alignment took place during <u>all or almost all</u> NLSA rounds. ¹³
Alignment with Other Assessments	The NLSA was <u>poorly aligned</u> with other types of assessment activities in the system.	The NLSA was <u>somewhat aligned</u> with other types of assessment activities in the system.	The NLSA was <u>very aligned</u> with other types of assessment activities in the system. ¹⁴	This option does not apply to this indicator.
Opportunities for Students to Be Exposed to Content and Skills	Students <u>did not have</u> <u>opportunities</u> to be exposed to the content and skills measured by the NLSA.	Students had <u>limited opportunities</u> to be exposed to the content and skills measured by the NLSA.	Students had <u>sufficient opportunities</u> to be exposed to the content and skills measured by the NLSA.	Students had <u>many opportunities</u> to be exposed to the content and skills measured by the NLSA. ¹⁵

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Preparatory Information for Schools	Official information on the NLSA was <u>not made available</u> to schools in the system.	A <u>minimal amount of official information</u> on the NLSA was made available to schools in the system, although not necessarily all schools. ¹⁶	A <u>sufficient amount of official information</u> on the NLSA was made available to most or almost all schools in the system.	An <u>extensive amount of official information</u> on the NLSA was made available to all or almost all schools in the system.
Quality Assurance	No <u>formal procedures</u> were in place to ensure the quality of the NLSA.	Formal <u>procedures</u> to ensure the quality of the NLSA were <u>minimal in nature or not required</u> .	Formal <u>procedures</u> to ensure the quality of the NLSA were <u>sufficient in nature and required</u> . ¹⁷	Formal <u>procedures</u> to ensure the quality of the NLSA were <u>extensive in nature and required</u> .
Standardization	The NLSA was <u>not standardized</u> at the system level.	The NLSA was <u>partially standardized</u> at the system level, or <u>minimal or no procedures</u> were in place to ensure standardization.	The NLSA was <u>fully standardized</u> at the system level, and <u>sufficient procedures</u> were in place to ensure standardization. ¹⁸	The NLSA was <u>fully standardized</u> at the system level, and <u>extensive procedures</u> were in place to ensure standardization. ¹⁸
Representativeness	A <u>nonrandom sample</u> or a <u>convenience sample</u> of students participated in the NLSA.	A <u>random sample</u> of students that was <u>not representative</u> at the country level participated in the NLSA.	All students in public schools, or a <u>representative sample</u> of students in public schools, participated in the NLSA. ¹⁹	All students in public and private schools, or a <u>representative sample of students</u> in public and private schools, participated in the NLSA.
Reasons for Not Taking the NLSA	<u>All or almost all individuals</u> could not take the NLSA due to <u>one or more non-assessment-relevant</u> reason(s).	<u>Most or some individuals</u> could not take the NLSA due to <u>one or more non-assessment-relevant</u> reason(s).	There were <u>no non-assessment-relevant reasons</u> that prevented individuals from taking the NLSA. ²⁰	This option does not apply to this indicator.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Quality Processes	<u>Many errors or delays</u> in activities took place that affected the NLSA to a <u>great extent</u> .	<u>Errors or delays</u> in activities affected the NLSA to a <u>significant level</u> .	Any errors or delays in activities had <u>only a minimal effect</u> on the NLSA.	<u>Errors or delays</u> in activities <u>did not affect</u> the NLSA. ²¹
Inappropriate Behavior	Inappropriate behavior <u>compromised the credibility</u> of the NLSA to a <u>great extent</u> .	Inappropriate behavior took place and <u>compromised the credibility</u> of the NLSA somewhat.	Inappropriate behavior was <u>low</u> and <u>did not compromise the credibility</u> of the NLSA.	Inappropriate behavior, if any, was <u>marginal</u> and <u>did not compromise the credibility</u> of the NLSA. ²²
Methods and Procedures Documentation	There was <u>no documentation</u> on the methods and procedures used during the NLSA.	There was <u>minimal documentation</u> on the methods and procedures used during the NLSA, or the documentation that existed was not public. ²³	There was <u>sufficient and public documentation</u> on the methods and procedures used during the NLSA.	There was <u>extensive and public documentation</u> on the methods and procedures used during the NLSA.
Publication Results	NLSA results were <u>not published</u> .	<u>Limited information</u> on the NLSA results was published, or the results were published using a <u>minimum number</u> of dissemination mechanisms.	<u>Sufficient information</u> on the NLSA results was published using an array of dissemination mechanisms. ²⁴	<u>Comprehensive information</u> on the NLSA results was published using an array of dissemination mechanisms.
Credibility of Results	The results of the NLSA were <u>perceived as credible</u> by <u>very few</u> stakeholder groups.	The results of the NLSA were <u>perceived as credible</u> by <u>some</u> stakeholder groups.	The results of the NLSA were <u>perceived as credible</u> by <u>most</u> stakeholder groups. ²⁵	The results of the NLSA were <u>perceived as credible</u> by <u>all or almost all</u> stakeholder groups.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Impact Monitoring	No mechanisms were in place to monitor the impact of the NLSA.	Minimal mechanisms were in place to monitor the consequences of the NLSA, or the mechanisms took place in <u>only</u> some or a few NLSA rounds. ²⁶	Sufficient mechanisms were in place to monitor the impact of the NLSA, and the mechanisms took place in <u>all or almost all</u> NLSA rounds.	Extensive mechanisms were in place to monitor the impact of the NLSA, and the mechanisms took place in <u>all or almost all</u> NLSA rounds.
Readiness to Start an NLSA Program	The system was <u>weakly prepared</u> to start an NLSA program in the future.	The system was <u>somewhat</u> prepared to start an NLSA program in the future.	The system was <u>well prepared</u> to start an NLSA program in the future.	This option does not apply to this indicator. ²⁷

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

1. The Republican Monitoring of students' learning achievement in school subjects was first implemented in 2003, and until 2011 it covered grades 4, 9, and 11. In the 2012–13 and 2013–14 school years, it covered grades 3, 8, and 10. Between 2003 and 2014, students' knowledge and skills were assessed in the subjects of mathematics and Belarusian and Russian languages. The assessment also covered biology, chemistry and physics in grades 9 and 11 (in 2003–11) and in grades 8 and 10 (in 2012–14). In 2012 and 2013, knowledge and skills in social studies were assessed in grades 9 and 10. In the 2014–15 school year, students' knowledge and skills were assessed in the following subjects and grades: grade 6: mathematics, world history; grade 7: physics, biology; grades 5 and 8: reading. From 2003 to 2006, Republican Monitoring was regulated on an annual basis by the Minister of Education of the Republic of Belarus, and since 2007, it has been regulated by the Program on the Development of General Secondary Education for 2007–16.
2. The policy-mandated purposes of this assessment include informing policy and pedagogy, as well as monitoring education quality and evaluating interventions aimed at improving student learning. This assessment is also used to evaluate the extent to which student learning is aligned with the national education standards and syllabi requirements.
3. The Belarus secondary education development program for 2007–16 (Council of Ministers of the Republic of Belarus, 2007) is the formal/official publicly available policy document that authorizes the Republican Monitoring.
4. The decree “On Organization and Carrying Out Monitoring in Secondary Educational Institutions” (Ministry of Education of the Republic of Belarus) provided guidelines on the Republican Monitoring, including on its governance, frequency of administration, who should be assessed, what should be assessed, how the results should be used, and how stakeholders should be informed of assessment results (including through publication of assessment materials in scientific-methodological subject journals, in the magazine *Vesnik Adukatsii*, and on the national education portal). Guidelines on the confidentiality of results are not provided by official documents.
5. The Department of Education Quality Monitoring was established at the National Institute of Education as a permanent unit with the responsibility for scientific-methodological and organizational support of the Republican Monitoring of Education Quality. Five assessment programs fall under the Republican Monitoring, including the assessment of:
 - a. Students' learning achievement in school subjects (the assessment program that is the subject of this report)
 - b. Student fatigue and efficiency (study of student functional status as well as factors such as the motivation to learn, the level of general learning abilities, academic load, compliance with sanitary requirements as to teaching arrangements, etc.)
 - c. Student personal development and “educatedness” (study of student perceptions of moral and legal norms in social life; student value system, behavior attitudes, level of student professional identity, student social adjustment degree, their awareness of help lines, etc.)
 - d. Quality of education services (study of organization efficiency of optional classes, industry-oriented teaching, extra school day, family-school interaction, etc.)

- e. Teachers' professional competency (study of teachers' proficiency, their educational orientation) is conducted by the Department of Education Sociology and Continuing Education of the Academy of Postdiploma Education.

The Department of Education Quality Monitoring of the National Institute of Education has had the primary responsibility of overseeing key national large-scale assessment scientific-methodological and organizational activities as related to the Republican Monitoring since 2003. The unit is responsible for assessment design, administration, data processing, and data reporting, including the preparation of data analysis and reporting of assessment results as well as for providing recommendations on the basis of the assessment results.

The Ministry of Education oversees the project management activities of the assessment and ensures the quality of the assessment materials.

In accordance with the Order of the Ministry of Education of the Republic of Belarus, committees that are responsible for education management in an oblast (1) select the coordinator for conducting the Republican Monitoring in institutions of general secondary education; (2) organize the participation of external observers (specialists from the rayon/city educational and methodological departments, specialists from the institutes of educational development) and the carrying out of the Republican Monitoring in regional education institutions; (3) organize technical support for the implementation of the Republican Monitoring in education institutions in the region (copying of materials, scanning of materials, carrying out a survey on the computer); and (4) provide verification and assessment of student responses in educational institutions and in the institutes of education quality. School coordinators, who are appointed by the director of an educational institution, are responsible for the implementation of assessment activities in the educational institutions. Temporary subject commissions, comprising subject teachers with the highest qualifications and teacher-methodologists (at the rayon level and in the institutes of educational development), are formed for the purposes of verification and evaluation of students' work.

Various departments and regional authorities, including those overseeing education at the oblast level and city (for example, the committee on education of the city of Minsk) and departments of education, sport, and tourism, conduct an external review of carrying out the assessment in the education institutions.

- 6. The Department of Education Quality Monitoring is part of the National Institute of Education, which is accountable to the Ministry of Education.
- The Ministry of Education is accountable to the Council of Ministers of the Republic of Belarus, and the Department of Education with oblast (rayon) executive committees is accountable to the oblast (rayon) executive committee.
- No unit with assessment responsibility is accountable to an autonomous board or committee that is institutionally separate from units in charge of the NLSA.
- 7. The budget for Republican Monitoring was provided by national and local government funding sources and covered assessment design, administration, data processing, data reporting, and project management. Research and development activities on the topic of assessment of the quality of education are also financed through other government sources.

8. Funding was sufficient to cover all core assessment activities.
9. The NLSA unit has most of the appropriate resources. Appropriate storage facilities and communication tools were the most readily available resources, while computers for all technical staff, software, and servers as well as building security were generally widely available.
10. Individuals responsible for completing key assessment activities included the permanent staff from the Department of Education Quality Monitoring of the National Institute of Education, permanent staff from the Research and Development Centre of the National Institute of Education (who carry out research on assessment issues, perform external review of diagnostic materials for monitoring, and develop educational standards and syllabi), experts from local educational authorities or local institutions of education development are involved as external observers, and teachers (teacher-methodologists of the highest qualification who work on subject commissions). All of the individuals responsible for completing key assessment activities had relevant qualifications for their roles.
11. Individuals responsible for completing key assessment activities included permanent staff from the Department of Education Quality Monitoring of the National Institute of Education, permanent staff from the Research and Development Centre of the National Institute of Education, and education specialists hired to support the assessment cycle. Generally no issues were found with the effectiveness of these individuals. Subject-matter teachers of the highest qualification are also invited to work on subject commissions. Experts from the Department of Education Quality Monitoring developed scoring guides to facilitate teachers' review and evaluation of students' tests. The subject committees include only top-ranked teachers and supervising teachers. Some issues were found with the effectiveness of these teachers (for example, some teachers overlooked errors, which was noted by the Department of Education Quality Monitoring of the National Institute of Education) specialists; however, the quality of the assessment overall was not compromised as a result.
12. There were sufficient opportunities to learn about the NLSA. Opportunities included:
 - a. University graduate programs (master's or doctorate level) on student assessment that include topics relevant to the NLSA (e.g., test design, reporting). "Monitoring of the quality of knowledge on the subject" courses are organized in higher pedagogical educational institutions.
 - b. University courses/workshops on the content and skills measured by the NLSA (e.g., courses on curriculum). In higher pedagogical educational institutions, as part of pedagogy and methods of teaching courses, topics include subject syllabus and learning outcomes to be measured in the course of education (as well as monitoring). For example, "Monitoring of the Quality of Knowledge on the Subject" courses are made available.
 - c. Nonuniversity courses/workshops on the content and skills measured by the Republican Monitoring (e.g., courses on curriculum). In continuing education institutions (Academy of Postdiploma Education, regional education development institutions) advanced training course syllabi cover topics such as learning outcomes and measurement approaches. For example, the National Institute of Education carries out live and distance-learning courses and methodological workshops for subject teachers on the topic of "Teacher and Student Control and Assessment."

- d. Nonuniversity courses/workshops on Republican Monitoring topics other than the content and skills measured by the assessment (e.g., test design, reporting). The Republican Institute of High School (RIHS) runs advanced training courses on the “Basic Theory and Methods of pedagogical measurements.”
- e. Funding for attending international programs, courses, and workshops on student assessment that cover topics relevant to the Republican Monitoring Education. Quality assessment system experts participate in education quality assessment training courses (including carrying out monitoring research) in the Russian Training Centre of the Institute of Education of NRU “Higher School of Economics” and in the Eurasian Association of Educational Assessment (EAOKO) and participate in international expert consultations on student academic achievement assessment.
- f. Opportunities to participate in webinars, distant learning courses, and online student academic achievement conferences. Specifically, education experts participate in online student academic achievement conferences, in webinars of the Russian Training Centre of the Institute of Education of NRO “Higher School of Economics,” and benefit from the study materials published on the Eurasian Association of Educational Assessment (EAOKO) website.

Opportunity (a) is available only to those studying in higher education institutions. Opportunities (b), (c), and (e) are available only to existing full-time staff at the NLSA unit, university professors, and primary and secondary school teachers and educators. Opportunities (d) and (f) are available only to existing full-time staff of the NLSA unit, other specialists of the National Institute of Education, and specialists of RIKC.

University courses/workshops on NLSA topics other than the content and skills measured by the Republican Monitoring (e.g., test design, reporting), internships/short-term employment in the unit running the Republican Monitoring, or presentations about the Republican Monitoring (e.g., test design, administration) are not made available.

13. The Republican Monitoring is intended to measure official learning standards and the official curriculum, with which it is closely aligned. Officially mandated reviews ensure that the Republican Monitoring measures what it is intended to measure as part of the Republican Monitoring, including regular internal and external reviews that take place during all or almost all Republican Monitoring rounds. Specifically, diagnostic tools for the Republican Monitoring are subject to:
 - a. Internal review in the Department of Education Quality Monitoring of the National Institute of Education
 - b. Informal external review in laboratories of the Research and Development Centre of the National Institute of Education
 - c. Formal external review in the Ministry of Education of the Republic of Belarus (in accordance with the decrees of the Minister of Education, all diagnostic and instruction materials for carrying out the Republican Monitoring are to be coordinated in the Departments of the Ministry of Education).
14. Although the format of the questions across assessment types may differ, classroom assessment and examinations are closely aligned with the Republican Monitoring.

15. Public school educators provide students with the opportunity to be exposed to the content and skills measured by the Republican Monitoring.
- School educators provide all or almost all students with the opportunity to be exposed to the content and skills measured by the Republican Monitoring through regular course instruction at school.
- Students can also access samples of previously administered Republican Monitoring assessments.
- Students in private institutions of general secondary education comprise 0.07 percent of students in institutions of general secondary education in the Republic of Belarus and are therefore not included in the Republican Monitoring sample. Thus, although public schools are covered in this review, private schools are not.
16. During meetings with coordinators of the assessment in institutions of general secondary education in the regions, which take place on an annual basis at the National Institute of Education, information on the main areas of the assessment, the objectives and content of the assessment in the given year, and the types of questions that will be on the assessment are discussed. Similar meetings are held immediately before the implementation of the Republican Monitoring in every region for coordinators in those institutions where the Republican Monitoring will take place (those institutions that are included in the sample).
17. Sufficient formal quality assurance procedures were in place to ensure the quality of the Republican Monitoring. Procedures that were required to take place every assessment round include:
- a. All proctors or administrators were trained according to a protocol
 - b. There was a standardized manual for NLSA administrators
 - c. All answer sheets were numbered
 - d. There were external observers (e.g., specialists from the rayon or city teaching offices of institutes of educational development)
 - e. There were internal observers (e.g., NLSA staff observing at administration sites)
 - f. Double data scoring (conducted by rayon subject commissions and subject commissions formed on the basis of the oblast institute of educational development; in cases of discrepancies in scoring, National Institute of Education specialists conduct a review as well).
- Procedures that were not required to take place every assessment round include:
- a. Piloting of questions/items/tasks before the official NLSA administration
 - b. Training of scorers to ensure high interrater reliability
 - c. Double processing of data.
18. The Republican Monitoring is fully standardized at the system level. Assessment design, administration, scoring, and reporting are the same (or equivalent) for all students in the same assessment round. Specifically:
- a. Republican Monitoring papers and questions/items/tasks were the same or equivalent for all students

- b. Republican Monitoring administrators were trained to ensure that all students took the NLSA under the same conditions
- c. Quality control monitors/observers were used to ensure the same administration conditions in all locations where the NLSA was administered
- d. The same scoring criteria were used to correct the Republican Monitoring questions/items/tasks and
- e. Republican Monitoring results were computed using the same procedures for all students.

Republican Monitoring results are not reported to all students in the same way.

- 19. A representative random sample of public school students in the target grades is selected to participate in the Republican Monitoring.

Students in private institutions of general secondary education comprise 0.07 percent of students in institutions of general secondary education in the Republic of Belarus and are therefore not included in the Republican Monitoring sample.

- 20. The Republican Monitoring is administered to students studying in secondary schools under the secondary school curriculum. All eligible students can take part in the Republican Monitoring. Starting in 2014, students from all rural schools have an equal opportunity to be included into the Republican Monitoring sample. This required increasing the number of external observers by employing specialists from the rayon and city teaching offices of the institutes of education development.
- 21. Errors in item/question/task development and errors in scoring student responses did occur; however, these did not affect the Republican Monitoring.

To avoid incorrect instructional wording in diagnostic materials used on the Republican Monitoring, a decision was made to carry out pilot research to test the diagnostic materials before the administration of the assessment. Additionally, to avoid scoring errors, a decision was made to brief the subject committee members in charge of checking students' tests to provide additional guidance on the scoring procedures.

The following issues did not occur: errors in test design; delays in test design; delays in item/question/task development; errors in administering the NLSA; delays in administering the NLSA; poor training of NLSA administrators; delays in scoring student responses; errors in data processing; delays in data processing; errors in reporting results; delays in reporting results; results not being reported.

- 22. The only inappropriate behavior that took place was students copying from other students. This inappropriate behavior did not affect the credibility of the Republican Monitoring, and a decision has been made to increase the number of different versions of the assessment from two to four to address this issue.

The following inappropriate behaviors did not occur: leakage of the content of the NLSA paper or part of a paper before the NLSA administration; students being trained to answer specific questions on the NLSA; students being excluded from taking the NLSA; providing student assistance via teachers, supervisors, mobile phone, or equivalent during the administration of the NLSA; changing students' responses after they have submitted their NLSA papers; changing students' results (e.g., score points) after their papers have been scored.

23. Information on the evaluation and assessment of all questions is provided in the “Recommendations for Evaluation and Assessment of Students’ Work” document that is made available on the website of the management of the monitoring of the quality of education at the National Institute of Education on the day that the Republic Monitoring takes place. This information is made available to assessment coordinators in those educational institutions that took part in the monitoring. The following methods and procedures are not documented: construction of questions/items/tasks; pilot testing of questions/items/tasks; analysis of piloted questions/items/tasks; test assembly; reliability; scaling; setting cutoff scores.
24. An official report on the results of the Republican Monitoring is submitted to the Ministry of Education of the Republic of Belarus (in accordance with the decree of the Minister of Education). Based on the report, recommendations on enhancing the quality of secondary school education are prepared and published on the national educational portal and in research and methodological journals. The report provides information about the number of students who acquired the content of the taught (and assessed) subject at a high, sufficient, average, acceptable, and low level. Official results of the Republican Monitoring are made available at the Ministry of Education and in education-specific scientific-methodological journals, as well as on the national education portal, which publishes a summarized version with recommendations on how to improve the management of education quality. Results are communicated at the collegium of the Ministry of Education, at meetings of specialists of the education system, and as part of seminars for increasing teacher qualifications. National results at the question/item level are also available through the newspaper *Nastauitskaya Gazeta*. Teachers of general secondary education, parents, and the local authorities have access to this information.
25. Most stakeholder groups perceive the Republican Monitoring results as credible.
26. The impact of the Republican Monitoring on the education system is studied by experts of the Department of Education Quality Monitoring of the National Education Institute through:
- Focus-group studies involving secondary school teachers at the premises of local education development institutes (at advanced education courses, methodological workshops, and meetings) (which occurs during most NLSA rounds) as well as individual and group discussions with secondary school teachers at the later monitoring rounds (which occurs during most NLSA rounds)
 - Benchmarking of student academic achievement level for school subjects of various target groups (based on the results of various monitoring rounds) (which occurs during some NLSA rounds)
 - Content analysis of republished syllabi and textbooks (which occurs during some NLSA rounds).
- The following mechanisms are not in place to monitor the impact of the NLSA: oversight committee; expert review groups; funding for independent research on the NLSA; studies (e.g., effect on school practices) on the NLSA.
- Over the past six assessment rounds, Republican Monitoring results have informed:
- Syllabi adjustment for school subject upon republishing in 2009 and 2012

- b. Textbook editing by separate school subjects upon republishing, development of various teaching materials for separate subjects (e.g., biology)
- c. Decision making on introduction of prespecialized training at the stage II of secondary education (in the 2015–16 academic year)
- d. Adjustment of teachers' advanced training syllabi
- e. Annual development of work plans of subject teachers' methodological association.

27. This indicator does not apply to this rubric.

BELARUS

International Large-Scale Assessment (ILSA)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Stability of Participation	The system <u>did not participate</u> in an ILSA round in the last 10 years. ¹	The system <u>participated</u> in an ILSA round in the last 10 years, <u>but did not complete</u> it.	The system <u>completed one</u> ILSA round in the last 10 years.	The system <u>completed two or more</u> ILSA rounds in the last 10 years.
Policy Document	No <u>policy document</u> authorized the ILSA program.	An <u>informal/draft policy document</u> authorized the ILSA program.	A <u>formal/official policy document</u> authorized the ILSA program, but the document was <u>not available</u> to the general public.	A <u>formal/official policy document</u> authorized the ILSA program and was <u>available</u> to the general public.
Stability of Organization	There was <u>no unit</u> with primary responsibility for running the ILSA program.	There was <u>a unit(s)</u> with primary responsibility for running the ILSA program, but the unit(s) was <u>temporary or had been in place for less than 5 years</u> .	There was a <u>permanent unit(s)</u> with primary responsibility for running the ILSA program that <u>had been in place for 5 or more years</u> .	This option does not apply to this indicator.
Accountability of Organization	There was <u>no unit</u> with primary responsibility for running the ILSA program, or else the unit responsible was <u>not accountable</u> to a clearly recognized body.	The <u>unit(s)</u> with primary responsibility for running the ILSA program was <u>accountable to a clearly recognized body within the ILSA unit</u> .	The <u>unit(s)</u> with primary responsibility for running the ILSA program was <u>accountable to a clearly recognized body within the same institution</u> as the ILSA unit.	The <u>unit(s)</u> with primary responsibility for running the ILSA program was <u>accountable to a clearly recognized external body</u> .
Source of Funding	There was <u>no funding</u> available for ILSA activities.	The source of <u>funding</u> for the majority of ILSA activities was <u>loans, credits, grants, or equivalent</u> .	The source of <u>funding</u> for the majority of ILSA activities was the <u>government's internal funding sources</u> .	This option does not apply to this indicator.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Activities Funded	There was <u>no funding</u> available for ILSA activities.	Funding was <u>not sufficient</u> to cover all core ILSA activities.	Funding was <u>sufficient</u> to cover all core ILSA activities.	This option does not apply to this indicator.
Organization Resources	The ILSA unit <u>did not have</u> the appropriate resources.	The ILSA unit <u>had some</u> of the appropriate resources.	The ILSA unit <u>had most</u> of the appropriate resources.	The ILSA unit <u>had all</u> of the appropriate resources.
Qualifications of Staff	There were <u>no individuals</u> responsible for completing key ILSA activities.	<u>Some</u> of the individuals responsible for completing key ILSA activities <u>had the relevant qualifications</u> .	<u>Most</u> of the individuals responsible for completing key ILSA activities <u>had the relevant qualifications</u> .	<u>All or almost all</u> of the individuals responsible for completing key ILSA activities <u>had the relevant qualifications</u> .
Effectiveness of Staff	There were <u>no individuals</u> responsible for completing key ILSA activities.	The <u>responsible individuals</u> completed key ILSA activities, but there were <u>significant issues</u> in how these activities were completed.	The <u>responsible individuals</u> completed key ILSA activities, with <u>only some issues</u> in how these activities were completed.	The <u>responsible individuals</u> completed key ILSA activities and there were <u>no issues</u> in how these activities were completed.
Staff/Teacher Opportunity to Learn	There were <u>no opportunities</u> to learn about the ILSA.	<u>Opportunities</u> to learn about the ILSA were <u>minimal, or not of high quality, or did not benefit all key stakeholder groups</u> .	There were <u>sufficient high-quality opportunities</u> to learn about the ILSA that were available to <u>key stakeholder groups</u> .	<u>Opportunities</u> to learn about the ILSA were <u>extensive, of high quality, and benefited key stakeholder groups</u> .

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Alignment with Other Assessments	The ILSA was <u>poorly aligned</u> with other types of assessment activities in the system.	The ILSA was <u>somewhat aligned</u> with other types of assessment activities in the system.	The ILSA was <u>very aligned</u> with other types of assessment activities in the system.	This option does not apply to this indicator.
Opportunities for Students to Be Exposed to Content and Skills	Students <u>did not</u> have opportunities to be exposed to the content and skills measured by the ILSA.	Students had <u>limited opportunities</u> to be exposed to the content and skills measured by the ILSA.	Students had <u>sufficient opportunities</u> to be exposed to the content and skills measured by the ILSA.	Students had <u>many opportunities</u> to be exposed to the content and skills measured by the ILSA.
Quality Processes	Many errors or delays in activities took place that affected the ILSA to a <u>great extent</u> .	Errors or <u>delays</u> in activities affected the ILSA to a <u>significant level</u> .	Any errors or <u>delays</u> in activities had <u>only a minimal effect</u> on the ILSA.	Errors or <u>delays</u> in activities <u>did not affect</u> the ILSA.
Inappropriate Behavior	Inappropriate <u>behavior</u> compromised the <u>credibility</u> of the ILSA to a <u>great extent</u> .	Inappropriate behavior took place and <u>compromised the credibility</u> of the ILSA <u>somewhat</u> .	Inappropriate behavior was <u>low</u> and <u>did not compromise the credibility</u> of the ILSA.	Inappropriate behavior, if any, was <u>marginal</u> and <u>did not compromise the credibility</u> of the ILSA.
Meeting Standards for Publication	ILSA results for the system <u>did not</u> meet the <u>standards</u> required for publication in the international report.	ILSA results for the system met <u>sufficient standards</u> to be presented <u>beneath the main displays</u> in the international report.	ILSA results for the system met <u>all</u> of the <u>standards</u> required to be presented in the <u>main displays</u> of the international report.	This option does not apply to this indicator.

(CONTINUED)

	<p>ILSA results were <u>not published</u> in the system.</p>	<p><u>Limited information</u> on the ILSA results was published in the system, or the results were published using a <u>minimum number of mechanisms</u>.</p>	<p><u>Sufficient information</u> on the ILSA results was published in the system using an <u>array of mechanisms</u>.</p>	<p><u>Comprehensive information</u> on the ILSA results was published in the system using an <u>array of mechanisms</u>.</p>
<p>Publication of Results</p>	<p>The results of the ILSA were perceived as <u>credible by very few</u> stakeholder groups.</p>	<p>The results of the ILSA were perceived as <u>credible by some</u> stakeholder groups.</p>	<p>The results of the ILSA were perceived as <u>credible by most</u> stakeholder groups.</p>	<p>The results of the ILSA were perceived as <u>credible by all or almost all</u> stakeholder groups.</p>
<p>Credibility of Results</p>	<p>ILSA results were <u>not used</u> by stakeholders in the system.</p>	<p>ILSA results were <u>used in minimal ways</u> by stakeholders in the system.</p>	<p>ILSA results were <u>used in sufficient ways</u> by stakeholders in the system.</p>	<p>ILSA results were <u>used in extensive ways</u> by stakeholders in the system.</p>
<p>Use of Results</p>	<p>The system was <u>weakly prepared</u> to participate in an ILSA program in the future.</p>	<p>The system was <u>somewhat prepared</u> to participate in an ILSA program in the future.</p>	<p>The system was <u>well prepared</u> to participate in an ILSA program in the future.²</p>	<p>This option does not apply to this indicator.</p>
<p>Readiness to Participate in an ILSA</p>				

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

1. Belarus has not participated in an international large-scale assessment exercise.
2. Belarus is planning to take part in the 2018 Program for International Student Assessment (PISA 2018). Funding support for PISA 2018–related activities in Belarus will be provided through a World Bank project loan.

The Republican Institute for Knowledge Control (RIKC) will be responsible for PISA 2018 activities in the country. RIKC has prior experience carrying out assessments because its main function is to carry out the Centralized Testing in Belarus. Specialists who work at the National Institute of Education, the Academy of Post-diploma Education, institutes of educational development, and local educational authorities are also prepared to work on PISA 2018, mainly because of their previous experience working on the Republican Secondary Education Quality monitoring program. These experts have degrees in pedagogy, social studies, and psychology, and many of them have completed training courses relevant to PISA, including a “Basic Theory and Methods of Pedagogical Measurements” course. National education quality assessment system experts regularly participate in webinars conducted by the Higher School of Economics Russian Training Centre at the Institute of Education in Moscow and receive information on education quality assessment from Eurasian Association for Educational Assessment (EAOKO) news bulletins.

Experts from the National Institute of Education also have experience conducting research in the area of educational assessment. In 2013 the institute, in association with the Russian National Training Foundation, conducted research to measure the information and communication competencies of grade 9 students in Belarus. The “Information and Communication Competence Assessment of Grade 9 Students in the Republic of Belarus” project was implemented within the framework of a small grants program implemented by the Russian Centre for International Cooperation in Education Development (CICED).

Acknowledgments

This report was prepared by the World Bank SABER–Student Assessment Team (Marguerite Clarke, World Bank Senior Education Specialist; Julia Liberman, World Bank Operations Officer; and Rachel Cooper, World Bank Education Consultant) in collaboration with Igor Kheyfets, World Bank Senior Economist and Task Team Leader for education projects in Belarus. The Team is grateful for the feedback and support from Irina Oleinik, World Bank Operations Officer; Hanna Shvanok, World Bank Team Assistant; and Anna Sakalouskaya, World Bank Program Assistant. This report was made possible with the input and contributions of Mr. Valeriy Alexandrovich Struzhko, Director of the Minsk Secondary School no. 161, who oversaw the data collection for the SABER–Student Assessment exercise in Belarus. The Team would like to thank Ms. Valentina Vasilevna Ginchuk, Head of the Department of Education Quality Monitoring at the National Institute of Education, and Mr. Viachaslav Zairovich Suleymanov for their invaluable inputs, feedback, and collaboration.

References

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

UNESCO Institute for Statistics. Belarus Education Statistics.
<http://www.uis.unesco.org/Pages/default.aspx/>.

World Bank. 2015. Belarus Country Indicator Data. World Bank, Washington, DC:
<http://databank.worldbank.org/data>.

———. 2015. “Belarus Country Overview.”
<http://www.worldbank.org/en/country/belarus/overview>.

———. 2015. Belarus—Education Modernization Project. Project Appraisal Document. World Bank, Washington, DC.

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

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

This work is a product of the staff of The World Bank with external contributions. The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

