Serbia

STUDENT ASSESSMENT

Key Policy Areas for Student Assessment

Status

1. Classroom Assessment

Emerging

There are several formal policy documents that provide detailed guidelines on classroom assessment. Although these policy documents are available online, and required to be available in primary and secondary schools, teachers are usually familiar only with those provisions that relate to the number of assessments that are required to be carried out. There are limited required uses of classroom assessment information to support student learning. In general, classroom assessment practices provide little feedback to students and parents, tend not to be aligned with the pedagogical or curricular framework, and are considered weak.

2. Examinations

Emerging

The Final Exam in Primary School Education was administered for the first time in 2011 to grade 8 students. According to the Law on the Foundation of the Education System, the three main purposes of the Final Exam are: (i) student certification for primary school completion, (ii) student selection to secondary school, and (iii) monitoring of education quality levels. There are limited systematic mechanisms in place to ensure the quality of the Final Exam, and there are limited options for students who do not perform well on the exam. No mechanisms are in place to monitor the consequences of the Final Exam.

3. National Large-Scale Assessment (NLSA)

Emerging

The National Assessment of Fourth Grade Students was administered in 2006 to a sample of students. The main objective of the survey was to collect data on education practices and overall levels of student achievement in Serbian language and mathematics for the purpose of improving the education system. Limited resources did not allow for all student groups to be represented in the sample chosen for the national assessment. While Serbian policymakers support national assessment exercises in principle, there is no plan for, or budget allocated to, future national large-scale assessment exercises.

4. International Large-Scale Assessment (ILSA)

Established

Serbia has participated in a number of international assessment exercises, including TIMSS (2003, 2007, 2011) and PISA (2003, 2006, 2009). Serbia will participate in PISA 2012, and is planning to participate in PISA 2015 and TIMSS 2015. However, there is no formal policy document that addresses the country's participation in international assessments. Funding for Serbia's participation in international assessment exercises, including the recently concluded PISA 2009, has been provided by donors, including the World Bank.

Systems Approach for Better Education Results

THE WORLD BANK

Introduction

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

What is SABER-Student Assessment?

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

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

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

SABER-Student Assessment methodology

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

Assessment types and purposes

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

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

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

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

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

Quality drivers of an assessment system

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

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

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

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

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

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

Source: World Bank.

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

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

Levels of development

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

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

The basic structure of the rubrics for evaluating data collected using the standardized questionnaires is summarized in Appendix 2. The goal of the rubrics is to provide a country with some sense of the development level of its assessment activities compared to best or recommended practice in each area. For each

indicator, the rubric displays four development levels—*Latent*, *Emerging*, *Established*, and *Advanced*. These levels are artificially constructed categories chosen to represent key stages on the underlying continuum for each indicator. Each level is accompanied by a description of what performance on the indicator looks like at that level.

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

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

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

Education in Serbia

Serbia is an upper-middle-income country in South Eastern Europe. GDP per capita (current US$) is $6310.40, with annual growth of 2 percent in 2011.

The Serbian education system is comprised of pre-primary (for children between the ages of 0.5 and 6.5 years); primary education from grades 1 through 8; and secondary education, which consists of either four years gymnasium or three to four years of vocational. Primary education is compulsory, and the preparatory pre-primary grade has been mandatory since the 2006-2007 school year. The gross enrollment rate in primary education is approximately 95 percent (in 2011) and the completion rate is approximately 94 percent. The gross enrollment rate in secondary education is approximately 91 percent and the completion rate is approximately 99 percent.

Serbia has made important efforts to improve its education system, particularly given the government's goal of acceding to the European Union. The government's priorities for education include adoption of new approaches to classroom teaching and enhancement of the integrity and efficiency of education administration through development of education management information systems.

Detailed information was collected on Serbia's student assessment system using the SABER-Student Assessment questionnaires and rubrics. It is important to remember that these tools primarily focus on benchmarking a country's policies and arrangements for assessment activities at the system or macro level. Additional data would need to be collected to determine actual, on-the-ground practices in Serbia, 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 Serbia's immediate interests and needs. Detailed, completed rubrics for each assessment type in Serbia are provided in Appendix 5.

Classroom Assessment

Level of development

Emerging
●●○○

The *Law on the Fundamentals of the Education System (2009),* the *Regulations on Student Assessment in Primary Education (2011),* and the *Regulations on Student Assessment in Secondary Education (1999)* provide detailed guidelines on classroom assessment. Although they are available online, and required to be available in primary and secondary schools, teachers usually are familiar only with those provisions that relate to the number of assessments that are required to be carried out.

There are a variety of system-wide resources available to teachers for classroom assessment. Education standards describe what students are required to know and be able to do at the end of the two cycles of primary education as well as at the end of four years of secondary education. Teachers are also provided with textbooks or workbooks that assist them with implementing classroom assessment as well as with scoring criteria/rubrics for evaluating students' work.

There are some system-level mechanisms to ensure that teachers develop appropriate skills in classroom assessment. For example, teachers are required to receive 120 hours of accredited training over five years and may elect to receive some of this on classroom assessment. Additionally, the Institute for Education Quality and Evaluation implemented one-day trainings for about 12,000 teachers, principals, and school counselors in 2010. Training topics have included implementation of education standards and student assessment. However, there are limited systematic mechanisms to monitor the quality of classroom assessment practices. Classroom assessment is a required component of a teacher's performance evaluation and is a required component of school inspection or teacher supervision.

Several recently-adopted rulebooks describe the relationship between assessment and the competencies of teachers. Educational advisors and school inspectors monitor and evaluate teachers' performance with respect to classroom assessment, taking into consideration these rulebooks.

Classroom assessment information is required to be disseminated to students and parents. Parents and students receive information on the student's performance at the end of the semester and school year in an official form (student book). If a student has a negative assessment, the parent receives official information at the end of the first or third trimester as well. During the academic year, parents receive information on student performance during regular parent-teacher meetings.

There are limited required uses of classroom assessment information to support student learning. These include providing feedback to students on their learning, informing parents about their child's learning, and planning next steps in instruction. In reality, 'on-the-ground' classroom assessment practices provide little feedback to students and parents, are often not aligned with the pedagogical or curricular framework, and are considered weak.

Suggested policy options:

1. Introduce required courses for teachers on classroom assessment, especially on the topic of formative assessment. Additionally, develop minimum standards for teacher competencies in classroom assessment and make the minimum standards part of the core curriculum for pre-service teacher training. For teachers already in service, introduce professional development training on classroom assessment.

2. Introduce mechanisms to monitor the quality of classroom assessment practices into the mandate of educational supervisors. This will help ensure that classroom assessment is based on standards, and is consistent and comparable between schools.

3. Develop regular school reports that combine the results of the external examination and classroom assessment to enable schools to better monitor the quality of their classroom assessment activities.

Examinations

Level of development	Emerging ●●○○

In Serbia, the Final Exam in Primary School Education was administered for the first time in 2011 to grade 8 students. According to the *Law on the Foundation of the Education System*, which authorizes the examination, its three main purposes are for (i) student certification of primary school completion, (ii) student selection to secondary school, and (iii) monitoring of education quality levels.

Secondary schools enroll students using the "total enrollment score" (school grades, which account for 60 percent + Final Exam score, which accounts for 40 percent). Students can apply to various schools with their school grades and Final Exam scores and select which school to attend from the ones that accept them. Students who do not perform well on the examination may choose to enroll in less selective secondary schools. If a student is not satisfied with their Final Exam results and cannot enroll in a preferred secondary school, he or she can re-take the Final Exam in June of the following year.

The material needed to prepare for the Final Exam is widely accessible by over 90 percent of the students in a variety of learning contexts (e.g., covered in public school, available for home schooling, available on line). Materials include examples of the types of questions that are on the exam, information on how to prepare for the exam, and a framework document explaining what is measured by the exam.

There have been coordinated attempts to improve the Final Exam by stakeholder groups, which have been welcomed by the Ministry of Education and Science and the Institute for Education Quality and Evaluation (the Institute). Both entities share the responsibilities for carrying out the Final Exam. The Institute prepares the exam framework, test materials, data analysis, and the data report. The Ministry of Education and Science is responsible for the implementation of the Final Exam, including the distribution of test materials, organizing exam rooms, test administration, exam supervision, test marking, and data entry. The Institute is accountable to the Ministry of Education and Science, which is accountable to the Government of Serbia.

Regular funding for the Final Exam is provided by the government and covers all core examination activities (examination design, administration, data processing, and reporting), as well as research. Development activities have been supported under the European Union *Support for Quality Assurance Within the National Primary and Secondary Education Examination System* project.

Teachers are involved in most examination-related tasks. The Institute establishes working groups of subject teachers for writing items and developing the examination scoring guide. Teachers are also involved in scoring the examination. Additionally, teachers serve as test administrators and supervisors, but not in their own school. At the same time, there are no general courses or workshops on the Final Exam available to teachers. Teachers who are involved in examination activities only receive instructions and guidelines for the implementation of those activities.

There is some documentation about the technical aspects of the Final Exam, but it is not in a formal report format. Additionally, limited systematic mechanisms are in place, such as internal review of the quality of items and test materials, and external review of exam implementation, to ensure the quality of the Final Exam. There are also no mechanisms in place to monitor the consequences of the Final Exam.

Suggested policy options:

1. Develop and implement additional mechanisms for assuring the quality of invigilation and marking of the Final Exam, such as ensuring that all students take the exam under the same conditions, and that test marking is objective, reliable, and valid.

2. Develop and implement mechanisms for the improvement of the quality of the Final Exam items and instruments.

3. Establish a national commission for examinations composed of key stakeholders with a mandate to monitor the quality of the Final Exam.

National Large-Scale Assessment (NLSA)

Level of development

Emerging ●●○○

The National Assessment of Fourth Grade Students was administered for the first time in 2006. The test was administered to a representative sample of students educated in the Serbian language who were able to take paper-and-pencil tests. The main objective of the survey was to collect data on existing education practices and overall levels of student achievement in Serbian language and mathematics for the purpose of improving the education system.

While there are formal and informal policy documents that authorize the National Assessment of Fourth Grade Students – including the *Law on the Foundations of the Education System (2009)*, the *Law on Amendments and Additions to the Law on the Fundamentals of the Education System (2011)*, and the draft of the *Law on Primary Education* (2012) – these policy documents are available to the public. There is also no firm plan in place for carrying out the assessment in the coming years.

In Serbia, NLSA studies are financed on an ad hoc basis by donors due to the lack of financial resources in the budget of the Ministry of Education and Science. The funding provided covers all core NLSA activities (design, administration, analysis, and reporting), as well as staff training. Funding does not cover research and development activities. The limited resources available meant that not all student groups were represented in the sample chosen for the 2006 National Assessment of Fourth Grade Students.

The Institute for Education Quality and Evaluation (the Institute) is in charge of the National Assessment of Fourth Grade Students. While the knowledge and competence of Institute staff is sufficient, the staff has to work overtime when it needs to complete tasks for other assessment-related activities (such as the Final Exam in primary education).

There are some opportunities to prepare individuals for work on the National Assessment of Fourth Grade Students, including university graduate programs (masters or doctorate level) and courses (graduate and non-graduate) specifically focused on educational measurement and evaluation. In addition, internships and short-term employment opportunities at the Institute provide hands-on NLSA experience. There are no courses or workshops on national large-scale assessment available to teachers in the country.

Mechanisms are in place to ensure the quality of the National Assessment of Fourth Grade Students, including verification of the quality of the test items and instruments by external experts and psychometric analysis, and use of a standardized manual for the assessment administration.

While there is some documentation about the technical aspects of the National Assessment of Fourth Grade Students, it is not in a formal report format.

Results from the National Assessment of Fourth Grade Students are used by some stakeholder groups in a way that is consistent with the purposes and technical characteristics of the assessment. For example, the Institute used the 2006 assessment results as an input for developing standards of achievement for the first cycle of education as well as for developing criterion-referenced tests for schools. At the same time, there are no mechanisms in place to monitor the consequences of the National Assessment of Fourth Grade Students.

Suggested policy options:

1. Adopt a policy document prescribing how NLSA data will be used to support monitoring and improvement of the quality and equity of education.

2. Ensure that all student groups are included in the sample for the NLSA (minority languages and students with learning difficulties and/or disabilities).

3. Develop a multi-year plan for carrying out NLSA activities.

4. Ensure that NLSA activities are part of the national multi-year budget planning, and include resources to allow for the inclusion of vulnerable student populations in the national assessment.

International Large-Scale Assessment (ILSA)

Level of development

Established
●●●○

Serbia has participated in a number of international assessment exercises, including TIMSS (2003, 2007, 2011) and PISA (2003, 2006, 2009). Serbia will participate in PISA 2012, and is planning to participate in PISA 2015 and TIMSS 2015. However, there is no formal policy document that addresses the country's participation in international assessments.

Funding for Serbia's participation in international assessment exercises, including the recently concluded PISA 2009, has been provided by donors, including the World Bank. Funding for research and development related to these assessment exercises and their results has been provided by the Institute of Psychology at Belgrade University (which implements PISA) and by other institutes that request specific research and development studies.

There is a team and national coordinator to carry out the ILSA activities. For example, in addition to the National Project Manager, the team responsible for carrying out PISA consists of a core group of 4-5 people and a large number of associates who are included in specific phases of the study (administration, marking, data entry, etc.). The core staff is comprised of full time employees at the Institute of Psychology who have regular duties other than PISA activities. The team is sufficiently staffed, although the size and structure of the team has not always been aligned to the needs of the project. In general, the trend has been to increase the size of the core team. Due to financial constraints, the team was able to attend only some of the international workshops and meetings for PISA.

Opportunities to learn about international assessments are available to individuals working directly on a specific international assessment exercise, and to university students studying assessment or a related area. For example, a masters-level university course on educational assessment that includes the topic of international assessments is offered at the University of Belgrade.

To date, Serbia has met all technical standards required to have its data presented in the main displays of the international reports for each ILSA. The country also has contributed new knowledge on international large-scale assessment to the global evidence base.

While media coverage of ILSAs is limited to a few small articles, the discussion of (for example) PISA results in the Serbian media has tended to focus on key issues, including how PISA results can help improve education.

Results from the ILSAs are used in some ways to inform decision making in Serbia. For example, the results and analysis of PISA 2009 data has informed the recently announced National Strategy for Development of Education until 2020. Additionally, the results of PISA are used by the National Education Council for monitoring the quality and equity of education in Serbia. PISA results have been used to inform teacher training programs and other assessment activities in the system (e.g., classroom assessment, examinations), for developing education standards for the end of compulsory and secondary education, and for informing the framework for the law on education and regulations governing the assessment practices in schools.

Suggested policy options:

1. Prepare a policy document that provides guidelines for Serbia's participation in ILSAs, specifying the rationale for Serbia's participation in a select number of international large-scale assessments.

2. Increase local capacity on the use of ILSA databases to inform policy decisions.

3. Ensure that ILSA activities are part of the national multi-annual budget planning, and include the cost of participation in the chosen number of ILSAs, as well as the follow up analytical work to inform policy making.

Appendix 1: Assessment Types and Their Key Differences

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

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

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

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

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

Appendix 4: Methodology for Assigning Development Levels

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

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

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

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

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

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

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

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

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

Appendix 5: SABER-Student Assessment Rubrics for Serbia

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

SERBIA
Classroom Assessment

ENABLING CONTEXT AND SYSTEM ALIGNMENT

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

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

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

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

Classroom Assessment: Development-level rating justifications

1. Three official documents authorized by the Ministry of Education and Science, provide the guidelines on classroom assessment:

    a. The Law on the Fundamentals of the Education System (2009)
    b. Regulations on student assessment in primary education (2011)
    c. Regulations on student assessment in secondary education (1999)

2. The documents are available online and are required to be available in primary and secondary schools. In practice, however, legal documents are often at the principal's, pedagogue's or psychologist's office in schools and teachers do not always know the regulations well or are familiar with only the provisions that relate to the number of required assessments.

3. The resources that are typically available include a document that outlines what students are expected to learn and the level(s) of performance that students are expected to reach in different subject areas at different grade/age levels. In Serbia education standards of student achievement describe what students are required to know and be able to do at the end of two cycles of the primary education as well as the standards at the end of four years of secondary education. Teachers are also provided with textbooks or workbooks that assist them with implementing classroom assessment as well as scoring criteria/rubrics for students' work.

    After the final exam at the end of primary education, the Institute for Education Quality and Evaluation provides schools with a report on the achievements of their students. This report also includes comparison of school achievement in the Serbian language or mother tongue and mathematics, with the average achievement of the municipality, county, school authority and the Republic and comparison of class achievements in the Serbian language or mother tongue and mathematics.

    All schools receive the *Assessment from idea to practice* (Ministry of Education and Sports, 2003) handbook during the training for applying "descriptive assessment" (narrative report on the standards that students achieved, as opposed to just school marks on the 1 to 5 scale). The handbook consists *Assessment towards outcomes*, written by Dragica Pavlovic-Babic and associates, and *Assessment for student development* written by Nenad Havelka, Emina Hebib and Aleksandar Baucal.

4. The two official documents are the *Standards of student achievements at the end of the primary education* document and the S*tandards for the end of the first cycle of the primary education* document that addresses Serbian language, mathematics, nature, and society.

5. Teachers must receive 100 hours of accredited training every five years. The training can be selected from a list of accredited trainings. The Ministry can also require specific training to be offered to teachers in certain academic years, in which case teachers are required to participate in this training. For example, the Ministry of Education and Science has organized trainings for all   teachers teaching first grade, and principals and school counselors in primary schools. The Institute for Education Quality and Evaluation has implemented an in-service teacher training seminar on assessment for primary and secondary school teachers, principals and school counselors since 2005.

The Institute for Education Quality and Evaluation has implemented one-day trainings for about 12,000 teachers, principals and school counselors in 2010. Training topics included implementation educational standards in planning class work and evaluating students.

The Institute for Improving Education invites applications each year for professional development programs for school staff. Various institutions and organizations submit their teacher training programs for accreditation. Once a training program is accredited, teachers may choose to participate in it for their professional development. These programs are published each year in the catalogue which is available on-line and in a paper version.

Some of the courses are systematic and some are ad-hoc.

6. Although teachers rarely rely mainly on multiple-choice, selection-type questions and it is not common to observe errors in the scoring or grading of students' work, classroom assessment practices are mainly about recalling information, provide little useful feedback to students, and are mainly used as administrative or control tool rather than as pedagogical resource. Additionally, it is very common to observe uneven application of standards for grading students' work and grade inflation is a serious problem. It is common for teachers to not use explicit or a priori criteria for scoring or grading students' work and parents are often poorly informed about students' grades. It is also common for classroom assessment practices to not be aligned with the pedagogical or curricular framework.

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

    Several rulebooks that were recently adopted describe the relation between assessment and the competencies of teachers. Educational advisors and school inspectors monitor and evaluate teachers' performance with respect to classroom assessment taking into consideration these rulebooks. These rulebooks include:

    - The Regulation on continuous professional development and gaining the title of a teacher, preschool teacher or school councilor

    - The Regulation on the working license for teachers, preschool teachers and school councilors

    - The Regulations on professional-pedagogical monitoring,

    - The Regulations on evaluating the quality of educational institutions is in the final phase of development. It will regulate the external evaluation of schools that will be based on the Quality standards for educational institutions

8. Teachers are required to report on individual student's performance to students and parents. Parents and students receive information on student's performance at the end of the semester and school year in an official form (student book). If a student has a negative assessment, the parent receives official information at the end of the first or third trimester as well. During the academic year, parents receive information on student performance

18

informally during regular parent-teacher meetings. These meetings are held by the class-teacher or main subject-teacher. Parents can also obtain information on student's performance during individual consultations, usually on pre-scheduled dates. Parents can talk to the class-teacher or subject-teacher. While teachers should meet with parents during regular parent-teacher meetings, the frequency of these meetings is not strictly regulated.

9. The required uses of classroom assessment activities to promote and inform student learning include providing feedback to students on their learning, informing parents about their child's learning, and planning the next steps in instruction.

Regulation on assessing students was changed based on the changes in the education system. In more detail than before, the regulation describes the ways of assessing and the criteria that needs to be fulfilled for a particular assessment. It points out the need to respect the conditions in which the student learns, as well as the circumstances in which teaching takes place. The key innovations are also the obligatory initial testing by instruments created by teachers in order to assist in planning of teaching/learning. Formative classroom assessment is introduced in this way. The general criteria and the framework for assessing the quality of learning as well as the criteria for assessing student behavior are defined in this regulation. The innovations in the regulations refer to the relation between the assessment and education or achievement standards. Students with special needs or students in disadvantaged living conditions who have an individual educational plan (IEP) are evaluated based on the adjusted requirements and standards that are defined in that IEP. Under the Regulations, the assessment is supposed to be an objective and reliable measure of progress and development, and an indication of quality and efficiency of the work of teachers and schools.

SERBIA
Examinations

ENABLING CONTEXT

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

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

(CONTINUED)

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

(CONTINUED)

22

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

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

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

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

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

(CONTINUED)

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

Examinations: Development-level rating justifications

1. The Final Exam in Primary School Education is a new examination program that started in 2011. Students in grade 8 are required to take the exam. The Law on the Foundation of Education System defines the three main purposes of the exam: student certification for grade or school cycle completion; student selection to secondary school; and monitoring education quality levels. This law also identifies the secondary purposes of the examination: planning education policy reforms; school or educator accountability; and self evaluation of schools.

    While the Final Exam is taken for selection for enrollment in secondary schools, there is no formal cut-off score. Different secondary schools enroll students with different total enrollment score (school grades + Final exam score). For example, students applying for secondary medical schools need to earn more than 95 points out of 100 to be enrolled. Students can apply to various schools with their school grades and exam scores and then they select which school to attend out of the ones that accept them.

2. Policy documents that authorize the Final Exam include:

    - The Law on the Fundamentals of the Education System (2009), authorized by the Serbian Parliament.
    - The Law on Amendments and Additions to the Law on the Fundamentals of the Education System (2011), authorized by the Serbian Parliament
    - The Regulations on the Program of the Final Exam in Elementary Education (2011), authorized by the Serbian Government
    - Manual for Implementation of the Final Exam (2011), authorized by the Ministry of Education

3. The Laws, Regulations, and Manual are available online.

4. The Law on the Fundamentals of the Education System describes the purpose of the examination and specifies who can sit for the examination.

    The Regulations on the Program of the Final Exam in Elementary Education describe the purpose of the examination, outline the procedures for special/disadvantaged students, specify who can sit for the examination, explain the alignment of the examination with curricula and standards, and explain the format of the examination questions.

    The Manual for Implementation of the Final Exam outlines the governance, distribution of power, and responsibilities among key entities, outlines the procedures to investigate and address security breaches, cheating, or other forms of inappropriate behavior, and identifies rules about preparation.

    The policy documents do not describe the authorized uses of results or the sources of funding.

5. Policy makers strongly support the Final Exam, particularly because of their need to have an instrument for monitoring the quality and equity of primary education.

27

Teacher unions and educators support the introduction of the new Final Exam because it consists of items previously not available to students (which was not the case with the previously administered exam). Teacher unions and educators have expressed that cheating needs to be prevented through more effective test administration, and because students cannot fail (students are must take the Final Exam, but not to achieve certain minimal result), students may not be making a strong effort to prepare for the exam.

Students and parents accept the Final Exam because they are used to having an exam at the end of the primary education. They also believe that it is fair to make enrolment decisions based on results on a national exam together with school marks than only based on school marks.

The media follows the Final Exam only during its administration. The media provides information on the exam as opposed to criticism or praise.

The European Commission has provided strong technical and financial support for the development of the Final Exam within the IPA 2008 project Support for quality assurance within the national primary and secondary education examination system.

6. Efforts to improve the Final exam are coordinated by the EU project that provides support to the Ministry of Education for the improvement of the implementation of the Final Exam. The project also supports the efforts of the Institute for Education Quality and Evaluation to improve the quality of test materials, data analysis, and data reporting. The Institute for Education Quality and Evaluation ("the Institute") is also planning a consultation process with teachers and students to identify areas for potential improvement.

7. Efforts to improve the Final Exam are very welcomed by the Ministry of Education and Science and the Institute. For example, the Ministry negotiated with the European Commission to receive technical and financial support through the IPA project and the Institute initiated independently activities aimed toward improving its assessment activities.

8. The implementation of the Final Exam is supported by regular government budget. The budget covers preparation of test materials, data analysis and reporting by the Institute, as well as the administration of the exam by the Ministry. Additional activities related to the Final exam (for example, secondary analysis of the final exam results) are co-financed through IPA. The main aim of the IPA project is to build the capacity of the Institute.

9. Funding for core examination activities is provided by the government.

10. Developmental activities are supported by the EU project "Support for quality assurance within the national primary and secondary education examination system."

11. There is shared responsibility between the Ministry of Education and Science and the Institute for Education Quality and Evaluation for carrying out the Final Exam. The Institute prepares the exam framework, test materials, data analysis, and data report. The Ministry of Education and Science is responsible for implementation of the exam (distribution of test materials, organizing exam rooms, test administration, exam supervision, test marking, and data entry). The distribution of these responsibilities is identified in the Law on Foundations of Education System that was adopted in 2009.

12. The Institute is accountable to the Ministry of Education and Science, which is accountable to the Government.

13. Examination results are recognized by the certification or selection system in Serbia.

14. The Institute is accommodated in a secure building that has secure storage facilities and is equipped with computers. The Institute has access to computer servers, backup devices, and has adequate communication tools, however their quality and capacity are not aligned with the Institute needs – for example, current servers do not have enough capacity to support the Final Exam database, therefore the Institute must rely on external servers.

15. There is permanent or full-time staff, but it is insufficient to meet needs of the examination. Specifically, according to the strategy of the Institute as approved by the Government, the Institute should have 60 employees. However, because of budgetary restrictions the Institute has only 37 employees. For effective operation, at least 13 additional employees are needed.

    No systematic or serious issues have been identified with the performance of the human resources that are responsible for the Final exam. From time to time, however, some item turns out to be problematic and need to be excluded, marking criteria are applied somewhat differently across the country as teachers from different schools score the Final Exam papers, and/or some exam administrators do not strictly follow the rules and procedures for administrating the Final Exam. These issues, however, typically do not have serious impact on the quality and equity of the Final Exam.

16. The country offers university graduate programs (masters or doctorate level) and courses (graduate and non-graduate), for example at the University of Belgrade, that are specifically focused on educational measurement and evaluation. It is also possible for junior experts to do an internship at the Institute.

17. The Final Exam measures national school curriculum guidelines or standards. According to the Law on Foundation of Education System, the Final Exam is defined as an exam that serves to assess the level of attainment of general and specific standards of achievement in elementary school education.

18. The content of the Final Exam is defined by the Law on Foundation of Education System (general and specific standards). However, a key issue among educators is related to the content of items. Subject specialists argue for more items that assess the recall of key facts and knowledge from a given school subject, while education specialists argue for more items measuring subject specific and cross curricular competencies (items similar to those used in the PISA study).

19. The material needed to prepare for the Final Exam is widely accessible by all students (over 90 percent) in a variety of learning contexts (e.g., covered in public school, available for home schooling, available on line). Materials can be downloaded for free from the Internet site of the Institute. Materials include examples of the types of questions that are on the exam, information on how to prepare for the exam, and the framework document explaining what is measured on the exam, The framework and examples of items are published as a part of the Regulations on the Program of the Final Exam in Elementary Education, and information on how to prepare for the examination is published in the Manual for Implementation of the Final Exam.

While there is no provision for providing material to students who cannot access personal copies of the materials due to poverty or living in a distant rural area, during the month of May (prior to the administration of the Final Exam in June), 8th grade students have an opportunity to become familiar with the content and format of the exam by taking the trial version of the exam in Serbian language and mathematics.

20. There are no courses on the Final Exam for teachers. Teachers who are involved in certain examination activities only receive instructions and guidelines for the implementation of those activities.

21. The Institute establishes working groups of subject teachers for writing items and developing the scoring guide. Teachers are also involved in scoring the examination. Additionally, teachers serve as test administrators and supervisors, but not in their own school.

22. There is some documentation about the technical aspects of the Final Exam, but it is not in a formal report format.

    After implementation of the first Final Exam in 2011, the Institute prepared several analytical reports. These reports are only available to the audience for whom the reports were intended (schools or the Ministry).

    The report to the schools provided the comparison of a particular school in relation to the average achievement at the national, regional and municipality levels, by gender, subjects and items.

    The report for the Ministry of Education provided an analysis of Final Exam results by School Regional Administrations, region, municipality, and school level.

    The internal report on the quality of items and test instruments identified areas of improvement of the item writing and test development for future exams. The Institute is working on producing an additional report on the quality of scoring in order to identify potential issues and to develop instruments and procedures to improve marking.

23. Internal review (such as the internal review of the quality of items and test materials produced by the Institute), and external review (such as the supervision of exam implementation organized by the Ministry of Education and Science the (the implementation of the Final exam in 2011 was monitored in each school by one Ministry supervisor)), help ensure the quality of the Final Exam.

24. Copying from other candidates has been observed. In general, however, according to the Manual for Implementation of the Final Exam, schools are obligated to organize the Final Exam in one large room and to arrange the room in such way that the distance between students is at least 1.2 meters. Additionally, the administration of the Final Exam is supervised by an external supervisor who reports directly to the Ministry on the regularity of exam.

25. Although there is some indication that the administration of the Final Exam in some schools was not fully aligned with the rules, procedures, and guidelines set in the Manual for Implementation of the Final Exam, only few complaints were voiced by students and parents as these practices did not

have a significant impact on student chances to be enrolled in preferred secondary schools. The Ministry of Education and Science and the Institute are receiving support from the EU IPA 2008 project to improve the monitoring and quality assurance of the next Final Exam.

26. According to the Law on Foundation of Education System, all students are obligated to take the Final Exam at the end of grade 8. The Final Exam is free of charge for students therefore there are no financial barriers to take it. Additionally, students take the Final Exam in their own school and the exam is offered in the minority languages, which makes it accessible for all students. For students with special needs, the Final Exam can be delivered in different mediums (for example, Braille letters for students with vision impairment) or with accommodated content (for example, students with special needs receive test accommodations based on the personal Individual Education Program). In 2011, about 98.5 percent of students tool the Final exam in the first term (June 2011). The remaining 1.5 percent of students had an objective reason (for example, sickness) for not taking the Final Exam.

27. The results of the Final Exam are used only for placing students in secondary education programs and schools, and for monitoring of the quality and equity of the primary education by the Ministry of Education and Science. There is no systematic evidence of improper use of examination results.

28. Up to five days after the Final Exam, student results are publically displayed on a poster in the school.

29. Students who do not perform well on the examination may opt for less selective schools. Students take the Final exam at the end of June of the year when they complete grade 8. About 99 percent of students take the Final Exam in that term. If a student is unable to take at the end of June, he or she can do so in August of the same year. According to the Law on Foundation of Education System, students are obligated to only take the Final Exam, and not to achieve a certain minimum score. Therefore, students cannot fail the exam. However, because results from the Final Exam , together with school marks in Grade 6-8, constitute the total score (exam results 40 percent and school marks 60 percent) with which students apply for enrollment to secondary education programs and schools, students who do not perform well on the Final Exam can only be enrolled in a limited number of programs and schools. In the infrequent cases when a student is not satisfied with the exam results and cannot be enrolled in a preferred secondary school, he or she can take the Final Exam again the following June (the student can wait for the next Final Exam without repeating the grade, or can accept to be enrolled in some secondary school for which they have achieved the required entrance score). Students cannot repeat the grade or attend remedial/preparatory courses in order to prepare to retake the examination.

30. There are no mechanisms in place (such as funding for independent research on the impact of the examination, a permanent oversight committee, studies (e.g., predictive validity) that are updated regularly, regular focus groups or surveys of key stakeholders, and/or expert review groups) to monitor the consequences of the examination.

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

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

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

(CONTINUED)

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

(CONTINUED)

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

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

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

ASSESSMENT QUALITY

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

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

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

1. National Assessment of Fourth Grade Students is used to monitor the education quality at the system level, support schools and teachers, and for policy design, evaluation, and decision making. The assessment has been administered once in 2006 in Serbian Language (Reading, Writing, Grammar, Literature) and Mathematics (Natural numbers, Operations with natural numbers, Geometry, Fractions, Solving everyday life problems) to a representative sample of students educated in the Serbian language and who were able to take paper-and-pencil tests.

   The specific objectives of the National Assessment of Fourth Grade Students were to determine the knowledge, competences and skills in Serbian Language and Mathematics that students acquired by the end of the first cycle of education, to determine factors that affect student achievement, to evaluate knowledge and proficiency of educators, as well as school condition. The ultimate objective of the survey was to collect relevant data on existing education practices, and to learn about the level of student achievement and the factors that affect student achievement, for the purpose of improving the education system.

2. There are formal and informal policy documents that authorize the National Assessment of Fourth Grade Students:

   - The Law on the Foundations of the Education System (2009), authorized by the Parliament of the Republic of Serbia.

   - The Law on Amendments and Additions to the Law on the Fundamentals of the Education System (2011), authorized by the Parliament of the Republic of Serbia.

   - Draft Law on Primary Education (2012), authorized by the Parliament of the Republic of Serbia, which will regulate national testing. It is submitted to the Parliament of Serbia by the Ministry of Education, and can be downloaded from the Parliament of the Republic of Serbia website.

3. The policy documents are available to the public.

4. There is currently no large-scale assessment plan for the coming years or future assessment rounds.

5. Policymakers support national assessments in principle, yet due to the lack of financial resources, there is no annual program of national assessment studies in place.

   Teacher Unions have not addressed national assessments which may be due to the fact that there have only been two (Grade Three National Assessment of Student Achievement and National Assessment of Fourth Grade Students), neither of which has affected teachers.

   Because the media is generally interested in reporting on interesting results, they support the national assessment. However, the news that the media reports on the national assessment is not analytical or in-depth. The media has generally been more interested in international assessment studies, such as PISA.

Think thanks and NGOs support national assessment studies as they can use the results for analysis and reviews.

University staff strongly support the National Assessment of Fourth Grade Students because they understand the role and importance of the assessment for improving education quality. Nevertheless, among education specialists at the universities, there is concern that the assessment focuses only on outcomes of education, and that it may cause policy makers to neglect the process of education.

6. National assessment studies are financed on ad hoc basis by donors due to the lack of financial resources in the budget of the Ministry of Education.

7. Funding covers all core NLSA activities: design, administration, analysis and reporting.

8. Funding does cover staff training.

9. The Institute for Education Quality and Evaluation ("the Institute") is in charge of the National Assessment of Fourth Grade Students.

10. The Institute receives the request from the Ministry to carry out certain assessment studies and determines when those studies will be carried out. The Institute has the responsibility and autonomy to carry out those studies.

11. The Institute is accountable to the Ministry of Education. The Institute also employs expert consultants to complete work on specific assessment-related tasks. These experts are accountable to the Institute.

12. While the quality of the Institute staff is sufficient, staff has to work overtime when it needs to complete tasks for other assessment-related activities (such as for the Final Exam in Primary School Education, and others).

13. Opportunities to prepare individuals for work on the National Assessment of Fourth Grade Students that are available on an annual basis include university graduate programs (masters or doctorate level) and courses (graduate and non-graduate) specifically focused on educational measurement and evaluation. For example, the University of Belgrade offers the Educational Psychology program (at undergraduate and master levels) that includes a course Evaluation of Knowledge and Educational Outcomes, which is specifically focused on education assessments and examinations. This program has a long tradition, and the course that specifically focuses on educational assessment and examination was introduced in 2006. The quality of this program is considered good as the content of the program and the assessment course is up-to-date. The professors of the program have both good theoretical knowledge and practical experience, and provide up to date information on the knowledge and practice in the area of national assessment studies. Training workshops are held at international education conference, and conference attendees can attend these workshops. Additionally, it is possible to gain hands-on experience through internships or short-term employment in the large-scale assessment office.

14. The National Assessment of Fourth Grade Students measures performance against curriculum or learning standards.

39

15. What the National Assessment of Fourth Grade Students measures is largely accepted by stakeholder groups.

16. Mechanisms include regular internal review as well as regular independent review by qualified experts of the alignment between the assessment instrument and what it is supposed to measure. Specifically, the Institute engages external experts to evaluate the quality and validity of test items.

17. There are no courses or workshops on the National Assessment of Fourth Grade Students.

18. The Institute does not have sufficient resources to translate the National Assessment of Fourth Grade Students into minority languages.

19. Mechanisms to ensure the quality of the National Assessment of Fourth Grade Students include that the quality of items and test instruments are verified by external experts and are based on psychometric analysis, all proctors or administrators are trained according to a protocol, there is a standardized manual for large-scale assessment administrators, a pilot is conducted before the main data collection takes place, all booklets are numbered, there is double data scoring, and scorers are trained and monitored to ensure high interrater reliability, and data entry is monitored as well.

20. There is some documentation about the technical aspects of the National Assessment of Fourth Grade Students, but it is not in a formal report format.

21. Results of the National Assessment of Fourth Grade Students are disseminated within twelve months after the assessment is administered, reports with results are made available for all stakeholder groups, and the main reports on the results contain information on overall achievement levels and subgroups.

    The main reports on the results do not, however, contain information on trends over time overall and for subgroups, do not contain standard errors (measures of uncertainty), there are no media briefings organized to discuss the results, there are no workshops or presentations for stakeholders on the results, and the results are not featured in the media (in newspapers, magazines, radio, or television).

22. Assessment information is used by the Institute as an input for developing standards of achievements for the first cycle of education as well as for developing criterion-referenced tests for schools.

23. There are no mechanisms in place to monitor the consequences of the National Assessment of Fourth Grade Students.

SERBIA
International Large-Scale Assessment (ILSA)

ENABLING CONTEXT

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

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

(CONTINUED)

SYSTEMS APPROACH FOR BETTER EDUCATION RESULTS

LATENT	EMERGING	ESTABLISHED	ADVANCED
●○○○	●●○○	●●●○	●●●●
ENABLING CONTEXT 3: *Having effective human resources for ILSA*			
There is no team or national/system coordinator to carry out the ILSA activities.	There is a team or national/system coordinator to carry out the ILSA activities.	There is a team and national/system coordinator to carry out the ILSA activities. [7] ✱	This option does not apply to this dimension.
This option does not apply to this dimension.	The national/system coordinator or other designated team member may not be fluent in the language of the assessment.	The national/system coordinator is fluent in the language of the assessment. [8] ✱	This option does not apply to this dimension.
This option does not apply to this dimension.	The ILSA office is inadequately staffed or trained to carry out the assessment effectively.	The ILSA office is adequately staffed or trained to carry out the ILSA effectively, with minimal issues. [9] ✱	The ILSA office is adequately staffed and trained to carry out the ILSA effectively, with no issues.

SYSTEMS APPROACH FOR BETTER EDUCATION RESULTS

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

LATENT	EMERGING	ESTABLISHED	ADVANCED
●○○○	●●○○	●●●○	●●●●
SYSTEM ALIGNMENT 1: *Providing opportunities to learn about ILSA*			
The ILSA team has not attended international workshops or meetings.	The ILSA team attended some international workshops or meetings. [10] ✳	The ILSA team attended all international workshops or meetings.	This option does not apply to this dimension.
The country/system offers no opportunities to learn about ILSA.	This option does not apply to this dimension.	The country/system offers some opportunities to learn about ILSA. [11] ✳	The country/system offers a wide range of opportunities to learn about ILSA.
This option does not apply to this dimension.	This option does not apply to this dimension.	Opportunities to learn about ILSA are available to the country's/system's ILSA team members only. [12] ✳	Opportunities to learn about ILSA are available to a wide audience, in addition to the country's/system's ILSA team members.

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

LATENT	EMERGING	ESTABLISHED	ADVANCED
●○○○	●●○○	●●●○	●●●●
ASSESSMENT QUALITY 1: *Ensuring the quality of ILSA*			
Data from the ILSA has not been published.	The country/system met sufficient standards to have its data presented beneath the main display of the international report or in an annex.	The country/system met all technical standards required to have its data presented in the main displays of the international report.	The country/system met all technical standards required to have its data presented in the main displays of the international report. [13] ✱
The country/system has not contributed new knowledge on ILSA.	This option does not apply to this dimension.	This option does not apply to this dimension.	The country/system has contributed new knowledge on ILSA. [14] ✱
ASSESSMENT QUALITY 2: *Ensuring effective uses of ILSA*			
If any, country/system-specific results and information are not disseminated in the country/system.	Country/system-specific results and information are disseminated irregularly in the country/system.	Country/system-specific results and information are regularly disseminated in the country/system.	Country/system-specific results and information are regularly and widely disseminated in the country/system. [15] ✱
Products to provide feedback to schools and educators about the ILSA results are not made available.	This option does not apply to this dimension.	Products to provide feedback to schools and educators about the ILSA results are sometimes made available.	Products to provide feedback to schools and educators about ILSA results are systematically made available. [16] ✱
There is no media coverage of the ILSA results.	There is limited media coverage of the ILSA results.	There is some media coverage of the ILSA results. [17] ✱	There is wide media coverage of the ILSA results.
If any, country/system-specific results and information from the ILSA are not used to inform decision making in the country/system.	Results from the ILSA are used in a limited way to inform decision making in the country/system.	Results from the ILSA are used in some ways to inform decision making in the country/system. [18] ✱	Results from the ILSA are used in a variety of ways to inform decision making in the country/system.
It is not clear that decisions based on ILSA results have had a positive impact on students' achievement levels.	This option does not apply to this dimension.	This option does not apply to this dimension.	Decisions based on the ILSA results have had a positive impact on students' achievement levels. [19] ✱

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

1. Serbia has participated in a number of international large scale assessments, including TIMSS 2003, TIMSS 2007, TIMSS 2011, PISA 2003, PISA 2006, and PISA 2009. Serbia has consistently participated in PISA, and it most recently completed PISA 2009 which the remainder of the rubric refers to.

2. Serbia has taken concrete steps to participate in PISA 2012, and is planning to participate in PISA 2015 and TIMSS 2015.

3. There is no policy document that addresses participation in international assessments.

4. Funding for PISA 2009 is provided for by government funding that is supported by a World Bank loan. Funding for participation in other international large scale assessments is also provided for by government funding supported by loans (such as from the World Bank and the European Investment Bank).

5. Funding for PISA 2009 covered international participation fees, implementation of the assessment exercise in Serbia (e.g., printing booklets, travel to schools), attendance at international expert meetings for the assessment exercise, processing and analyzing data collected from implementation of the assessment exercise, and reporting and disseminating the assessment results in Serbia.

6. While donor funding does not cover research and development activities, the Institute of Psychology which implements PISA is a research institution. Additional research and secondary analyses are typically done by the Institute of Psychology and financed by the Institute itself by another organization at its request (e.g. Social Inclusion and Poverty Reduction unit of the Government of Serbia requested and financially supported the Institute of Psychology to make a targeted secondary analysis of PISA achievements and transition from compulsory to secondary education for students with lower SES. Results from this secondary analysis were used in the "First national report on poverty").

7. Dragica Pavlović-Babić is the National Project Manager for PISA. The team responsible for carrying out PISA consist of a core group (4-5 people) and a large number of associates who are included in specific phases of the study(administration, marking, data entry, etc.). The core staff are full time employees at the Institute of Psychology (i.e. formally at the Faculty of Philosophy) as teaching and research staff with regular duties other than PISA. The network of associates are not employees of the Institute but are brought in on a temporary basis as the majority of them are teachers employed by schools or Faculties/Universities. Some of the associates are students or unemployed persons brought in to complete data entry or other temporary administrative tasks. TIMSS, on the other hand, is administered by the Institute of Pedagogical Research, which is a public institute.

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

9. The team is sufficiently staffed, however the size and structure of the team has not always been aligned to the needs of the project. In general, the trend is to increase the size of the core team. Some members of the core team are stretched by different PISA tasks and other duties they have as full time employees of the Institute of Psychology and the Faculty of Philosophy. At the same time, the core team's composition is interchangeable because different members of the core team can take on various additional tasks, which is particularly helpful during the implementation of many PISA activities. The Institute can extend the core team at its own discretion, but the size of the PISA budget received from the Ministry limits this expansion. Lack of financial resources

keeps the number at a minimum. The team has previous experience working on international assessments as well as the necessary training to carry out the required PISA activities effectively.

The only issue identified with the carrying out of PISA in Serbia is that there have been errors or delays in the printing or layout of the test booklets. Specifically, the team in charge of the PISA exercise could not achieve the expected level of print quality (although the quality of the test booklets was still acceptable) due to lack of financial resources.

10. Team members attended most of the training organized by the OECD. Several trainings were not attended due to financial constraints (for example, if the same training was attended in previous cycles and if participation requires a high cost).

11. Limited opportunities to learn about international assessments include university courses on the topic of international assessments. A masters-level university course on educational assessment that includes international assessment studies is offered for students studying Education psychology at the University of Belgrade. The topic of the international studies will be also included in the new master program for educational policy at the University of Belgrade. Additionally, there is limited funding for attending international workshops or training on international assessments.

12. Individuals working directly on the specific international assessment exercise and university students studying assessment or a related area benefit from the opportunities to learn about international assessments.

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

14. Serbia has contributed new knowledge on PISA.

15. Results and information from the PISA exercise were regularly and widely disseminated in Serbia. Specifically, a national report was made available online, copies of the national and international report were distributed to key stakeholders, Serbia's results were communicated through a press release, brochures and PowerPoint presentations with Serbia's results were made available online or distributed to key stakeholders, and products providing feedback to the schools and educators about the results were made available.

16. Participating schools receive the school report as well as the National report (other schools can download the national report from Serbian PISA site). In addition, some schools have asked the National Center to visit the school and make a presentation and/or discuss PISA findings with teachers.

17. Media coverage has included editorials or columns commenting on the international assessment results, including, for example, how PISA results can help improve education. The quality of the discussion of international assessment results in the media has also become more professional. In 2012, media coverage of implementation of PISA was very high. For example, Politika, a reputable daily newspaper, published and analyzed the administered PISA items two times per week over three months.

18. The results and analysis of PISA 2009 data has informed the recently announced National Strategy for Development of Education until 2020. Additionally, the results of PISA are used by the National Education Council for monitoring the quality and equity of education in Serbia. PISA results were used in the First report on poverty reduction produced by the Government of Serbia. Additionally, PISA results have been used to inform teacher training programs and other assessment activities in the system (e.g., classroom assessment, examinations), for developing education standards for the end of compulsory and secondary education, and for informing the framework for the law on education and regulations governing the assessment practices in schools.

    The results of PISA 2009 have not been used to track the impact of reforms on student achievement levels, to inform curriculum improvement, or to inform resource allocation.

    Serbia is currently implementing a project of formulating education standards for the end of general secondary education. PISA results, as well as the PISA concept of literacy, informed the working groups responsible for developing the standards of students achievements.

19. There is evidence that the achievement of students in Serbia significantly increased in PISA 2009, in comparison to previous PISA cycles. However, during the same period, there were no systematic changes in the education system. At least in part, the growth of achievement could be explained by the changing teaching practices that resulted from specific teacher training programs, as well as by the increased general understanding of such concepts as education competence, quality of education, and functional literacy, all concepts which were informed by international assessments.

Acknowledgements

References

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

Organisation for Economic Co-operation and Development (OECD). 2012. "Strengthening Integrity and Fighting Corruption in Education: Serbia." Paris: OECD.
———. 2011. "Education at a Glance 2011: OECD Indicators." Paris: OECD.

United Nations Children's Fund (UNICEF). 2008. "Country profile: Education in Serbia." Data retrieved from http://www.unicef.org/ceecis/Serbia.pdfon March 12, 2013.

United Nations Educational, Scientific, and Cultural Organization (UNESCO). Serbia Country Indicator Data. Montreal, QC: UNESCO. Data retrieved from http://stats.uis.unesco.org/unesco on March 14, 2013

World Bank. 2013. "Serbia Country Overview" Data retrieved from http://www.worldbank.org/en/country/serbia/overview on March 16, 2013.

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

49

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

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

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

THE WORLD BANK