Disseminating and Using Student Assessment Information in Russia

Yulia Tyumeneva

THE WORLD BANK

Disseminating and Using Student Assessment Information in Russia

Yulia Tyumeneva

1818 H Street NW
Washington DC 20433
Telephone: 202-473-1000
Internet: www.worldbank.org

1 2 3 4 15 14 13 12

Cover design: Patricia Hord.Graphik Design, Alexandria, VA

Contents

Figures

Abstract

This case study examines how the Russian Federation disseminates and uses information from its student assessment system, drawing lessons for other countries seeking to more effectively use their own assessment data. Russia's Unified State Examination (USE) is primarily used for student selection and certification purposes, but a variety of other uses have been attached to it, including informing pedagogy, ensuring accountability, and monitoring education quality. This variety of uses has had both positive and negative consequences for the school system. Information from international large-scale assessments (e.g., PISA, TIMSS, and PIRLS)[1] has been widely used to introduce reforms in the school system. Factors that have affected the differential use of USE and international large-scale assessment data include, among others, the purpose of a given assessment, its design features and level of credibility, and access to the assessment database.

[1] PISA – Programme for International Student Assessment; TIMSS – Trends in International Mathematics and Science Study; PIRLS – Progress in International Reading Literacy Study.

About the Author

Yulia Tyumeneva heads the Educational and Psychological Measurement master's degree program at the Higher School of Economics, a national research university in Russia, where she is a lecturer and researcher. She is an expert in developmental psychology and educational assessment, with a deep knowledge of national and regional assessment systems. Her experience includes secondary analysis of national and international large-scale assessment results, as well designing questionnaires and surveys in the field of education. Tyumenva's current research project is entitled "Analyses of Factors Connected with Student Performance in International Comparison Research." She has a Ph.D. in psychology from the Russian Academy of Education.

Acknowledgments

For their advice and valuable comments, the author would like to thank Victor Bolotov, vice president of the Russian Academy of Education; Galina Kovaleva, head, Center for Assessment of Educational Quality, Russian Academy of Education, and national project manager for TIMSS, PIRLS, and PISA; and Isak Frumin, academic supervisor, Graduate School of Education, Higher School of Economics. Thanks also are extended to Sergey Bochenkov, head, Department for Assessment of Educational Quality, Chuvash Republic, for making available the data that were used in this paper.

Executive Summary

The purpose of this case study is twofold: (1) to present how the Russian Federation disseminates and uses student assessment information, and (2) to draw lessons for other countries aiming to improve the use of assessment information and, therefore, the effectiveness of their assessment systems. The paper examines two main types of assessments: the Unified State Examination (USE) of the Russian Federation and well-known international large-scale assessments.

The USE is the most important assessment program in Russia. The high-stakes national examination is used primarily to certify secondary school students and to select students into tertiary education. The USE is also used for a variety of other purposes, including informing pedagogy; holding regions, schools, and teachers accountable; and monitoring educational quality. While using the same instrument for a variety of purposes may seem efficient, it also can create problems. For example, since Russia does not have a national large-scale assessment program for system monitoring and accountability purposes, the USE has ended up being used to fill this gap, even though it is not actually designed to yield the kind of information required to carry out such system monitoring and accountability.

International large-scale assessments also play an important role in the Russian education system. Since the fall of the Soviet Union, Russia has consistently participated in the main international assessment programs, most notably, PISA, TIMSS, and PIRLS.[2] The results of these assessments have created awareness of the need to reform the school system in Russia. Access to the assessment databases has allowed for policy research and the formulation of action plans to introduce such reforms. International assessments have thus served as a reference for new national learning standards, curricula, and textbooks. There is a general consensus that the information generated by these assessments has been effectively used to improve educational quality in the country.

Several factors account for the differential impact of the USE and international assessments on the Russian education system. The USE was

[2] PISA – Programme for International Student Assessment; TIMSS – Trends in International Mathematics and Science Study; PIRLS – Progress in International Reading Literacy Study.

primarily designed to select and certify students, not to inform policy. The national examination program still needs to build its credibility and do a better job of refining and disseminating its conceptual framework as well as granting broader access to assessment data. In contrast, international assessments were designed with the primary purpose of informing policy, they have high credibility, and their data are widely available for policy research.

Countries aiming to improve the dissemination and use of the data generated by their assessment systems can learn from the following lessons. The effective use of assessment information has to be carefully planned; it will not happen spontaneously. Having clear purposes for the information is a starting point. Involving key institutions is required to meet those purposes. Disseminating information is a required, but insufficient, condition for success. The effective use of assessment data requires clear guidelines, action plans, and recommendations. Of note, key features of an assessment program affect the use of the data it generates. These features include, among others, a clear conceptual framework, reporting student performance levels based on learning standards, and attaching stakes to the assessment results. Reformers should be aware of these features in order to maximize the positive impacts of the assessment systems in their countries.

Disseminating and Using Student Assessment Information in Russia

Yulia Tyumeneva

Introduction

The effective use of student assessment information is key to improving a country's education system. Countries that have set up stable assessment systems often face many challenges in effectively disseminating and using assessment data. Russia's experience with secondary school examinations and international large-scale assessments provide useful insights into both the challenges and successes of the effective dissemination and use of assessment data. This paper can be useful for other countries that are introducing reforms to improve the impact of assessment information on educational quality and equity.

Currently, Russia administers the annual national Unified State Examination (USE) and regularly participates in international large-scale assessments (e.g., PISA, TIMSS, and PIRLS). The purposes of the USE and the international assessments differ significantly and have led to the results from these assessments having varying degrees of influence on the Russian education system.

The USE was introduced in 2009 with a dual purpose: to certify secondary school students, and to select students into tertiary education. This is a major examination program that has a great impact on student educational paths and careers. The USE is mandatory, taken every year by nearly a million students in 83 subject areas. The examination is fully standardized and centrally managed, thus allowing for greater quality, equity, and efficiency than institutional-level, unstandardized examinations.

Today, USE results are used for a myriad of purposes beyond student selection and certification. These include informing pedagogy; holding regions, schools, and teachers accountable; and monitoring educational quality. The high-stakes nature of the USE has led to a push

for a stronger focus on results in student learning. Yet it has also led to a push for the mechanical training of students on examination-taking techniques.

Participation in international assessments has been a key feature of Russian educational policy since the 1990s. After the fall of the Soviet Union, Russia, as a newly established federation, looked to revamp its education system in order to successfully insert itself into a global market-oriented economy. International assessments such as PISA, TIMSS, and PIRLS provided the conceptual basis and operational tools for reforming national learning standards, curricula, and textbooks. These reforms were further prompted by Russia's unsatisfactory results on the international assessments, particularly in the case of secondary school students.

There is a general consensus that international assessments have greatly contributed to improving education in Russia, but views differ regarding the contribution of the USE. Several factors have contributed to this differential effect. First, as already noted, the primary purpose of the USE is not to inform educational policy, whereas this *is* the primary purpose of international assessments. Consequently, the examination's ability to inform policy is limited. Second, access to USE data is restricted, and analysis of USE results, scarce. In contrast, access to international assessment data is open to the public and free; researchers publish analyses that directly inform policy. These analyses are then translated into recommendations and action plans.

Finally, use of USE results seems to be diverting the focus of many Russian educators from developing students' higher-order thinking skills to training them for the examination. International assessments, however, have pushed teachers to focus on higher-order thinking skills by emphasizing the application of knowledge to solve real-life problems.

This paper is organized as follows. The next section discusses the uses of student assessment information generated by the USE. Subsections discuss issues surrounding the use of USE results for certifying and selecting students into tertiary education, for informing pedagogy, for accountability purposes, and for system monitoring. The following section describes the uses of international large-scale assessment data to introduce educational reforms in Russia. The last section concludes with lessons learned regarding factors that impede or

facilitate the effective use of student assessment information to improve educational quality.

The Unified State Examination

The USE is Russia's most important assessment program. It was introduced at the national level in 2009 to replace and standardize a previously existing system of institutional-level final examinations. The USE was one of the core elements of a major reform of the education system in Russia. This was a very complex reform that required fighting corrupt practices in university entrance decisions and launching a massive public campaign to gain public support. Nowadays, nearly a million students take the examination every year, allowing them to apply to several tertiary education institutions through an automatic and objective process (Bolotov et al. 2013).

As already noted, the USE is used for a variety of purposes, which include secondary school certification; selection into tertiary education; informing pedagogy; holding regions, schools, and teachers accountable; and monitoring educational quality. This myriad of uses is in part the consequence of Russia not having a national large-scale assessment program for monitoring and accountability purposes. The USE is therefore used to meet functions for which it was not originally designed, with both positive and negative consequences.

Certifying and selecting students

Students receive USE scores based on their results in mandatory (Russian language and mathematics) and optional subject areas. These test scores can be used by students to apply to universities. The higher the test score, the higher the chance of getting into a more selective university. Students can apply for up to six universities and programs of study.

All universities require students to submit test scores in the mandatory subject areas and, depending on the program of study, optional subject areas. For example, on top of the mandatory subject areas, a medical program will require test scores in biology and chemistry; a technical program, in physics; and a program in the humanities, in history and social studies.

Test scores are also used to classify students into four academic performance levels: minimal, low, medium, and high. A minimal performance level does not qualify a student for high school certification. Cut-off points that define the performance levels are established each year by the government, based on that year's test results, official learning standards, and expert opinion. This post facto procedure avoids a situation in which too many students fail the examination. Unfortunately, the criteria used to establish the cut-off points vary from one year to the next, limiting the validity of yearly comparisons.

Examination results are produced and recorded by the Federal Testing Center of the Ministry of Education. This center sends individual test scores to municipalities and schools, which are ultimately responsible for reporting them to the students. Test results are free of charge and confidential. Each student receives a secondary school certificate with the transcript of his or her test scores.

Informing pedagogy

As a high-stakes examination, the USE greatly shapes teaching and learning in Russian secondary schools and beyond. USE results are used to create pedagogical guidelines to prepare future students for the examinations. These guidelines have come to serve as teaching aids for teachers and study guides for students. The examinations are expected to have a positive effect on educational quality because they are aligned with national learning standards. Given the stakes attached to the examination, both teachers and students are assumed to be highly motivated and focused on mastering the national learning standards (measured by USE).

Both federal and regional guidelines exist to prepare teachers and students for the USE. At the federal level, these guidelines include examination content for each subject area, sample questions, and criteria for scoring open-ended questions. At the regional level, guidelines alert teachers to current deficiencies in student readiness for graduation and provide guidance in overcoming these challenges. However, regional guidelines often focus on statistical analyses and neglect to translate these analyses into action plans for improvement. Consequently, recommendations made at the regional level are often simply duplications of federal-level recommendations.

Critics claim that the USE is having a detrimental effect on student learning and educational quality. Examination preparation—rather than teaching and learning high-order thinking skills—has become the key element of school action plans.[3] Although focusing on examination preparation runs counter to the advice of regional think tanks, this practice is facilitated by the easy-to-use guidelines for USE preparation. These guidelines have accordingly become a de facto curriculum for teachers, severely narrowing the content and cognitive skills taught in class. As a result, federal curricula guidelines and learning standards are sidelined. Schools with the highest potential risk of poor final test results (e.g., rural schools in small towns far from regional centers) seem to have the most accentuated focus on the use of examination guidelines.[4]

Schools with poor USE scores may be subjected to closer inspection. Inspectors gather school observations and conduct interviews with school directors and teachers in order to produce a diagnosis and suggest actions for improvement. For instance, inspections may show deficiencies in the organization of a school (e.g., students' failure to attend additional examination preparation lessons, an incomplete curriculum). Schools receive guidelines on how to address these weaknesses in order to improve student learning (and USE results). Unfortunately, these guidelines often miss the mark. They tend to narrowly conceptualize solutions, focusing on violations of school procedures and rectifying shortcomings in the implementation of a school's specific activities, such as improving low attendance in additional student preparation classes.

Some regional ministries establish funding priorities based on USE results. For example, a regional minister may announce the goal of raising USE scores for the region to the national level in natural sciences.[5] Accordingly, special funding is created in order to improve teaching and learning in this area.

A major weakness in the effective use of USE data is that teacher training programs do not seem to benefit from it. While these programs are provided with information about the examinations (Bolotov et al. 2013), they do not seem to use it to inform changes to their programs of

[3] See, for example, Municipal Institute of General Education of the Chuvash Republic (n.d.).
[4] Ibid.
[5] See, for example, Minister of Education and Science of the Krasnoyarsk Krai (Territory) (2010).

study, nor is it used as an input in class discussions. This is a serious misalignment that prevents future and in-service teachers from learning about examination results. One would expect that training programs would provide teachers with opportunities to learn about the content and skills measured by the tests, as well as to reflect on how that content and skills should be taught. However, if USE results are used to inform teacher training, care will need to be taken to prevent the mechanical use of USE guidelines as a de facto curriculum.

Holding regions, schools, and teachers accountable

USE results are used to put pressure on different levels of the school system by means of the school accreditation process, school rankings, the publication of school results, and the distribution of economic incentives to both teachers and schools.

With respect to school accreditation, schools are mandated by law to inform the federal Ministry of Education and Science of student results in compulsory subjects for the previous three years. In the case of persistently low scores, schools risk being refused renewed accreditation. While this scenario has yet to play out since the USE was implemented in 2009, the possibility remains.

Regional ministries of education and municipalities publish school rankings based on USE results. Most of the time, USE scores are the only criterion used to rank the schools, a practice that goes against the recommendations of both federal agencies and local think tanks. Measuring school quality based on the proportion of certified students is unfair to schools that work in disadvantaged contexts. These schools may be adding a lot of value to their students' education, but many of these students may still not reach the academic level required for certification. On the other hand, schools working in socially advantaged contexts may not be adding much value to their students' education, but may still produce a high proportion of certified students.

Schools compete to attract more students and funding by publicizing USE results. Thus, they often post on their websites reports that feature their average examination scores, the number of students meeting the minimum and maximum benchmarks, comparative analyses of school scores with those of other schools in the city or district, and their school ranking. Such reports are frequently used to promote a school's

performance in comparison to other schools. High USE scores greatly contribute to a school's reputation and are instrumental in attracting both additional funding from the government and new students to the school.

With respect to economic incentives, municipalities reward higher-performing schools based on their overall USE scores from the previous year. This approach suffers from two important weaknesses: (1) it does not take into account changes in USE results from one year to another, and (2) it does not take into account the socioeconomic background in which a school operates. Therefore, schools operating in socially disadvantaged backgrounds tend to be unfairly penalized when their USE scores are compared to those of schools operating in more socially advantaged backgrounds.

Examination results are also used to hold teachers accountable and to distribute bonuses. A new national wage system provides teachers with incentive bonuses based on the performance of their students on the USE. The amount of the incentive varies among schools and districts. This wage system is partly linked to a national teacher evaluation program that links student results to subject area teachers in grade 11. Examination results, moreover, have a strong impact on teachers' reputations; this has pushed teachers further in the direction of using USE preparation guidelines as a de facto curriculum.

Monitoring educational quality

USE results are a key indicator of educational quality at the national and regional level. Federal-level reports examine the extent to which students are meeting curricular standards in different subject areas. The Ministry of Education (through the Federal Institute of Pedagogical Measurement) publishes online aggregated information about student performance on the examinations. These data include average scores in each subject area, the percentage of students meeting different performance levels and passing the examinations, student performance on tasks that measure different content and skills, and item difficulty levels. Yearly comparisons are also reported, despite the fact that the tests are not equated and the criteria used to define the performance levels may vary from year to year.

Overall, the data are used to identify deficiencies in student knowledge and provide a foundation for action. For example, a USE report stated:[6]

> Exam-takers in 2009 as a whole showed low results in solving geometry problems, at both high and low levels of difficulty. Many exam-takers cannot solve geometry problems—not only advanced [problems], but also those at the basic level. These results reflect the situation in schools, which was clearly unfavorable for the study of geometry for many years, insofar as the State Final Certification [the examination in place before USE was introduced] only tested 10th and 11th grade algebra and data analysis.

Federal reports are discussed at conferences and technical meetings led by the Ministry of Education and Science and the Russian Academy of Education. Participants include specialists from federal and regional governments, as well as staff from education centers and teacher training institutes. Materials documenting key findings are distributed to participants, who then disseminate them in the regions.

Varied strategies are used to disseminate USE results at the regional level. Regional ministries and departments of education publish USE results on their websites. They also hold press conferences attended by the regional media. These press conferences usually focus on the average regional USE score by subject area, the number of graduates who passed the examination, and the number of students who achieved the minimum examination score. These figures are compared with nationwide results to contrast the regional performance with that of the rest of the country. USE results and analysis are also incorporated into regional annual reports on the status of education. Reports of excessive numbers of students failing to obtain secondary school certification raise concerns about the inefficient use of education budgets, as well as the ineffectiveness of regional governors.

By tracking changes in student performance from year to year, USE scores are also used to monitor trends in educational quality. For instance, changes in the percentage of students who meet the high performance level are used as an indicator of educational quality. However, this practice has been criticized because USE tests are not comparable over time. It is not possible to discern whether changes in test

[6] See Federal Institute of Pedagogical Measurement (2009: 69).

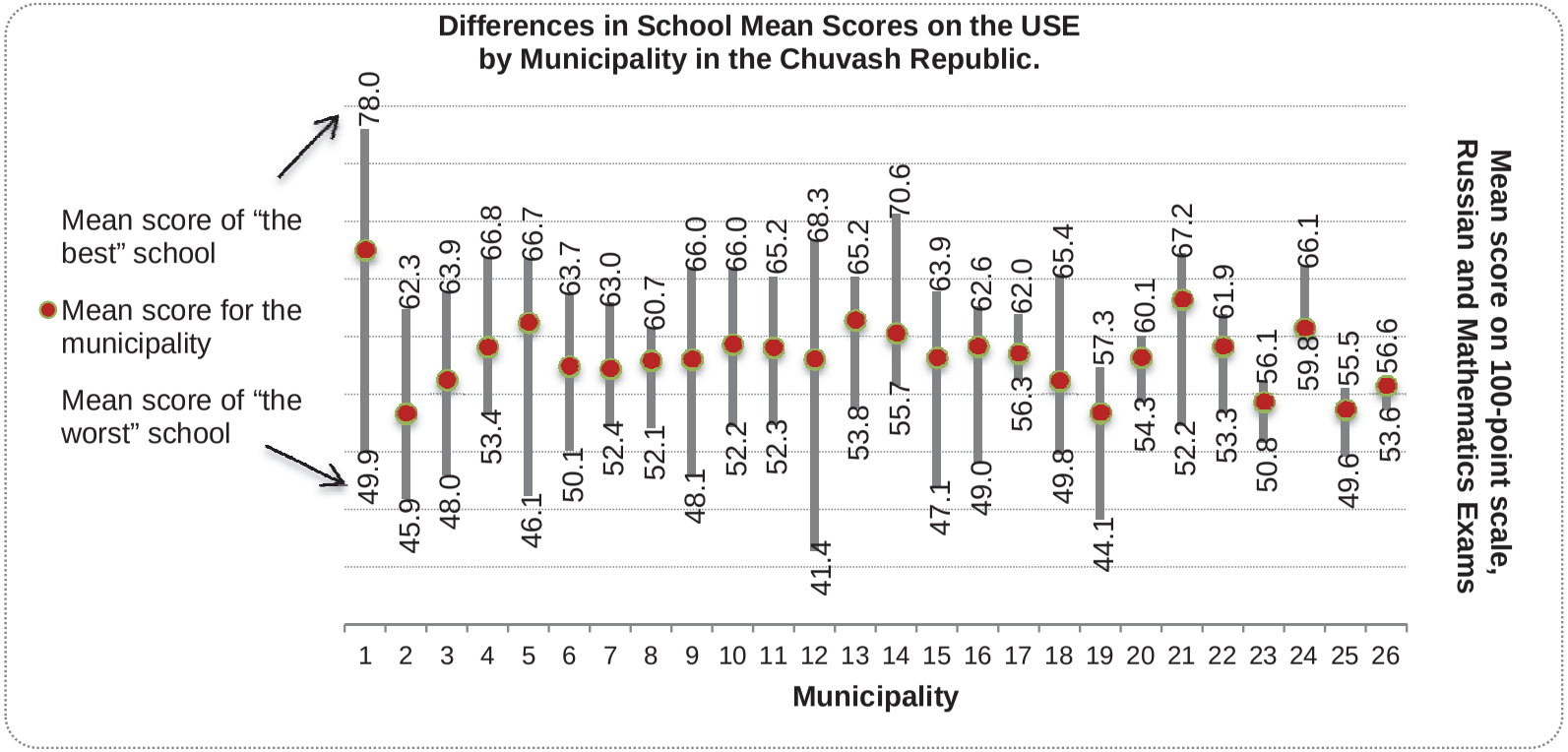
scores are due to changes in student performance or changes in the difficulty level of the tests administered in different years. To equate the tests, test scores would need to be rated on the same scale. This would allow for changes in test scores to be attributed to genuine changes in student performance. As noted above, comparing results from year to year has also been criticized due to variations in the way performance levels are defined.

Finally, USE results are used to monitor educational quality at the regional level. Regional centers independently analyze USE results, often producing comprehensive regional reports and analyses. For example, figure 1 shows the differences in the performance of schools in different municipalities of the Chuvash Republic. This type of analysis allows for an understanding of disparities in the quality of education within regions. It also provides local governments with the information necessary to target funding and provide additional support.

Nongovernmental institutions also analyze USE results to monitor educational quality and push for reforms. For example, the Higher School of Economics (through its Institute for Educational Development) discusses the results of USE research at weekly seminars. These discussions are often taken into consideration by the Committee on Education of the Public Chamber which, in turn, drives the adoption of initiatives by the Ministry of Education and Science. Unfortunately, analysis of student performance by individual schools is rare, largely due to the extremely limited number of specialists working in schools who are capable of analyzing such data.

While the USE provides a rich source of information for monitoring educational quality, the examinations also have important limitations. First, USE results only provide data on student performance at the end of secondary school, not their performance in earlier grades. Second, the examination differentiates between students who do and do not meet the national learning standards. It is not aimed at describing what students are able to do at different performance levels, which would be more appropriate for monitoring and pedagogical purposes. Since Russia does not have a national large-scale assessment program for such purposes, the USE seems to be filling this gap. As noted at the outset of this paper, this situation has both positive and negative consequences.

Figure 1. Regional Analysis of the 2011 Unified State Examination Results, Chuvash Republic



Source: Bochenkov 2011.

Note: The taller the bar, the greater the discrepancy between the best- and worst-performing schools. The red dots represent the average score of a given municipality.

International Large-Scale Assessments

After the collapse of the Soviet Union, there was a renewed interest in understanding the relative standing of Russian students compared to students in other Western economies. Were Russian students meeting international standards? Were they acquiring the skills needed to be successful citizens in a global market economy? What were students in other countries learning? Participation in international assessments allowed Russia to answer these questions. The assessments have provided rich information on what Russian students know and can do in key areas (e.g., reading, mathematics, sciences), compared to students in other countries. They have also identified trends in student performance by comparing the results of different assessment rounds. Finally, these assessments have provided Russia with rich contextual information that has been used to inform policy and introduce reforms.

As noted earlier, Russia has consistently participated in international large-scale assessments in the post-Soviet era. Since the collapse of the Soviet Union, it has participated in 19 such assessments, including TIMSS in grades 4, 8, and 11 (1995, 1999, 2003, 2007, 2008, 2011); PISA, which evaluates 15-year-old students (2000, 2003, 2006, 2009); PIRLS, which evaluates the reading literacy skills of grade 4 students (2001, 2006, 2011); and other similar instruments (CIVED, SITES).[7] These assessments are administered in Russia by a special unit of the Russian Academy of Education in collaboration with regional education centers (Bolotov et. al. 2013).

Findings from international assessments are widely disseminated and used to inform educational policy in Russia. Specifically, data are disseminated online, at conferences and seminars, and within regional education centers and teacher training institutes. This practice generates extensive discussions involving the Ministry of Education at the federal and regional level. The participation of curricular, textbook, and assessment specialists allows for the findings of international assessments to serve as a feedback mechanism for the education system. The outcome of these discussions are often considered at meetings of the Public

[7] CIVED – Civic Education Study; SITES – Second Information Technology in Education Study.

Chamber and then discussed at the regional and federal government level.

Mass media has been a major engine for disseminating international large-scale assessments results in Russia. The general public shows great confidence and interest in these assessments. In fact, there is a growing demand for information on comparative education. Between 2009 and 2011, for example, more than 300 print or electronic articles discussing PISA results were published, together with interviews with educational authorities.

As noted earlier, international large-scale assessments have had significant influence on education reforms in Russia. This influence has been largely triggered by concerns about educational quality and the country's ability to adapt to a global market economy. The performance of Russian students on TIMSS and PIRLS has always been relatively high compared to other developed countries. Concerns about educational quality in the country first arose after the PISA assessments of 2000 and 2003 when, for the first time, Russian students performed at levels below that of students from other developed countries. Russian secondary school students were losing the relative advantage they had demonstrated at the primary level. Moreover, results were not improving over time. These findings triggered profound reforms in the education system in order to align the Russian education system (including curriculum standards, textbooks, and examinations) with international teaching and learning trends.

Based on international assessment findings, the Russian Academy of Education has greatly supported the introduction of education reforms in Russia. This academy plays a triple role in the education system: (1) it is in charge of implementing international assessments in the country; (2) it is one of the main developers of national curricular standards; and (3) it provides guidelines and criteria for textbook development. This triple role facilitates the translation of international assessment findings into education reforms.

Reforms have also been possible thanks to policy research based on international assessment data. Three research centers in Russia are involved in the analysis of international assessments. The Russian Academy of Education (through its Institute of Content and Method of

Education)[8] produces the official national report for Russia, which is publicly available online and in hard copy. Secondary analyses are done by the Moscow School of Social and Economic Sciences and the Higher School of Economics. These analyses usually focus on factors associated with student performance, the difficulty level of test questions or problems, and trends in student performance.

International assessments thus led to changes in Russian learning standards and the basic education curriculum. Reforms to primary and secondary school curricula focused on new skills. At the primary level, curricular changes promote the solving of visual geometry problems, practical work with geometric objects, working with tables and diagrams, understanding the basic principles of probability theory and statistics, a greater focus on fractions and percentages, understanding numerical and symbolic sequences, and rounding and estimating calculation results.

At the secondary level, the new educational standards are based on PISA. More attention is now given to the practical application of knowledge, with a competency-based approach to learning introduced in the classroom. Curricular changes promoted the development of spatial concepts; the application of mathematical knowledge in real-life situations; problem solving; finding intermediate data; dealing with relationships, values, and percentages; and understanding the basic principles of probability and statistics.

Textbooks have been aligned with the content and skills measured by international assessments and designed to include more real-life examples and problems. They put greater emphasis, for example, on the application of conceptual models to solve practical problems. Textbooks are now also required to include interdisciplinary sections that combine reading, mathematics, and science topics and skills.

Finally, international assessments have greatly shaped the new USE, which largely adopted the PISA approach. This translated into a major emphasis on the use of student knowledge and skills to solve real-life problems and the presentation of problems that require knowledge and skills in different subject areas, among other changes. Within each subject examination, about a quarter of the problems assess students' abilities to apply their knowledge to new situations. In mathematics, for example,

[8] See the institute's website at www.centeroko.ru (accessed July 2013).

students are required to independently develop a model and method of solving a real-life mathematics problem. Students are then required to justify and show the steps taken to find the mathematical solution.

The USE also places greater emphasis on communication skills. For example, students are assessed on their ability to explain their point of view based on available information, evaluate written statements, and analyze the point of view of a source text. In a country that had no tradition of using multiple-choice questions, international assessments were critical to validating and introducing this new question format into the examinations.

Lessons Learned

Russia has taken big steps toward using student assessment information as a systematic feedback mechanism for the education system. Nevertheless, there are enormous differences in the way information from national examinations and international assessments are used.

Purpose of the assessment

The purpose of an assessment significantly influences the use of its findings. The USE, as noted, assesses student performance for the main purposes of secondary school certification and selection into tertiary education. In contrast, international assessments are designed to assess student knowledge and skills with the primary purpose of informing policy. They are also designed to make international comparisons, which draw media attention and put political pressure on education officials, especially in the case of poor results. It is therefore not surprising that international assessments have had a greater impact on the reform agenda in Russia.

Credibility

The credibility of the organizations leading an assessment, as well as that of the assessment instruments and procedures, is crucial for facilitating the use of assessment information. Without credibility, there is no political will to introduce reforms based on assessment findings. The USE examination still suffers from low credibility among members of the

public, education practitioners, and researchers. Although the federal government made great efforts to build trust in the examination, it still has a long way to go. A 2011 survey showed that only 20 percent of the country approved of the USE (Public Opinion Foundation 2011).

Building credibility requires publishing technical information about the examination and conducting regular validation studies. It also requires fighting corruption. Unfortunately, the corruption that the USE was intended to circumvent has found its way into the new assessment system. Scandals ranging from the leaking of examination papers ahead of time to the falsification of results have been reported. These practices have added greatly to the lack of public confidence in the USE.

In contrast, international assessments have high credibility among members of the public, education practitioners, and researchers. This credibility is based on the reputation of the organizations leading the assessments, as well as the proven quality of their instruments and procedures. Transparency in sharing technical information, together with rigorous quality controls, has been critical in building trust in international assessments. The different levels of credibility of the USE and international assessments explain to a great degree why the latter have had a greater impact on educational reforms in Russia.

Involvement of key institutions

The participation of key institutions in assessment activities is critical to promoting systemic reform. The triple role played by the Russian Academy of Education (responsible for implementing international assessments, developing curriculum standards, and providing textbook guidelines) has allowed for the conversion of international assessment findings into curriculum and textbooks changes.

From research findings to policy recommendations

Research findings are only useful when they are converted into policy recommendations. However, the Ministry of Education at the federal level does not have a team in place to analyze USE data and produce policy reports and recommendations. Consequently, USE findings have a limited impact on education policy. In contrast, research findings from PISA and TIMSS are generally accompanied by policy recommendations,

which have been widely adopted by the federal government in Russia and have significantly contributed to improving the education system.

Assessing the whole school cycle

Assessments that target different grades in the school cycle are more likely to be used for policy decisions. These assessments provide a more complete picture of the quality of the education system, allowing education policy makers to diagnose and rectify challenges at various stages of schooling. International assessments benefit from the fact that they target different grades and school cycles. For example, PIRLS assesses students at grade 4; TIMSS, at grades 4, 8, and 11; and PISA, 15-year-olds (typically, grade 10). The USE is limited to grade 11, when students finish secondary education. Therefore, there is a feeling that it is already too late to try to introduce improvements with respect to these students, since they have already left, or are about to leave, the school system.

Stakes attached to an assessment

The uses and impact of assessment information vary widely depending on the stakes attached to it. The USE is a high-stakes assessment that has direct consequences on student certification and selection into tertiary education. It also has a strong impact at the school level, with teachers preparing students and students studying for the examinations. These seem to be short-term strategies for taking the examination at the end of the school year.

As discussed earlier, the USE is also being used as an input into school accreditation and decisions on school and teacher bonuses. These additional stakes have created undue pressure for superior results, producing several negative consequences. Some teachers end up using the USE framework instead of the curriculum standards, which narrows the curriculum taught in class. Additionally, these incentives have led to reports of corruption in the administration of the USE, the leaking of examinations, and the falsification of results.

On the other hand, international assessments are low-stakes assessments that do not have any impact on students, their teachers, or schools (although they have a growing impact on holding the government accountable). These assessments have, however, had a

strong influence on educational reform in Russia, leading to long-term strategies for improving educational quality.

Assessment framework

Assessments should have well-articulated assessment frameworks that justify the assessment principles and approach, describe the knowledge and skills to be tested, and specify the types of problems and question format that will be presented to students. The strong framework of international assessments has provided a clear roadmap for establishing educational goals to guide curricular development, instruction, and assessment in Russia. For example, the PISA approach to assessment identifies key objectives that have been adopted into the new educational standards in Russia. These standards have since been incorporated into the design of the USE. These changes have created greater alignment between national education standards and examinations, on the one hand, and international assessments, on the other.

Performance levels

Problematic methods for determining student performance levels hamper the use of this information for pedagogical purposes. In Russia, USE performance levels are not defined based on national curriculum standards. Instead, they are defined based on political considerations. The methodology used by international assessments to define performance levels seems more technically sound and fair. These assessments typically report what students know and can do at different performance levels, and the procedures for defining these performance levels is more stable over time. As a consequence, the performance levels used in international assessments have been a rich source of information for Russian educators.

Comparisons of assessment results over time

A key indicator of educational quality is how far students progress from one assessment round to the next. This requires comparing student results from different assessment rounds and, usually, different test versions. To do this in a valid way, an assessment must meet certain technical characteristics, which usually involve putting questions from different tests on the same scale (i.e., equating). As already noted, USE

instruments are not equated from year to year. Therefore, it is impossible to tell if changes in examination results are due to "real" changes in student performance, changes in the difficulty level of the test, or changes in the cut-off points used to define performance levels. This uncertainty severely limits the validity of USE data for monitoring trends and holding schools accountable. The Russian federal government warns against making conclusions based on direct yearly comparisons. Nevertheless, regional ministries of education frequently report yearly comparisons, leading to inappropriate conclusions about changes in student performance. In contrast, international assessments use technically sound techniques that allow for the equating of tests from different years. This allows for valid yearly comparisons of test results.

Background information

Effective use of assessment data is greatly conditioned on the availability of student, teacher, and school background information. This information is critical to understanding the factors related to student performance, whether superior or inferior. International assessments have greatly benefited from the collection of a vast array of background information. For instance, TIMSS includes a survey of students, their teachers, and schools in order to identify factors affecting the quality of education. This assessment also collects detailed information about curriculum standards and their implementation in the classroom. This information is then used by the countries to inform decision making on education policy. In contrast, the USE does not collect background information. To date, the examinations only collect information on student gender and school location. Meaningful analyses would require knowing the socioeconomic status of students, schools, and teachers, among other data points.

Access to assessment database

Limited access to the USE database severely restricts the country's capacity to make use of these rich data to inform decision making. In Russia, only the Ministry of Education at the federal level has full access to these data. For reasons of confidentiality, very restricted information is circulated beyond this organization; this information, moreover, does not seem sufficient for meaningful secondary analysis. This means that other branches of government, independent researchers, and education offices

are unable to use the data to inform their practices. In 2011, however, the Committee on Education of the Public Chamber initiated discussions about possible limited access to a depersonalized USE database. By contrast, open access to international assessment databases enables independent researchers to contribute unique analyses to the research and policy agenda in Russia. These efforts have complemented the research done by the Russian Academy of Education and have compensated for federal agencies' limited resources to conduct further research.

* * * * * * * *

Countries aiming to improve the dissemination and use of student assessment information should be aware of the factors discussed in this section when designing their assessment systems. By using the right combination of factors in the design, they will be in a better position to meet the ultimate purpose of a student assessment system: to improve student learning and education quality.

Bibliography

Bochenkov, S., ed. 2011. "Analysis of Results of USE 2011 in the Chuvash Republic" (in Russian). Ministry of Education and Youth Policy of the Chuvash Republic, Cheboskary. http://ege21.ru/index/0-2 (accessed July 2013). Cited with permission of the ministry.

———. 2012. "Analysis and Interpretation of USE 2012 Results" (in Russian). *Management in Education: Theory and Practice* [Russian Academy of Education] 3: 9–34. http://www.iuorao.ru/20120930/216------2012

———. 2013. "Teacher, School, Education System through the Prism of USE" (in Russian). *Problems of Modern Education* 3: 27–47.

Bolotov, V., G. Kovaleva, M. Pinskaya, and I. Valdman, 2013. "Developing the Enabling Context for Student Assessment in Russia." SABER Student Assessment Working Paper 9. World Bank, Washington, DC.

Center for Quality Education Assessment at INCM RAE. Various years. Analytical reports on PISA, TIMSS, and PIRLS (in Russian). Center for Quality Education Assessment, Moscow. http://www.centeroko.ru/public.htm.

Clarke, M. 2012 "What Matters Most for Student Assessment Systems: A Framework Paper." SABER Student Assessment Working Paper 1. World Bank, Washington, DC.

Derkachev, P., and I. Suvorova. "Unified State Examination as a Way to Assess the Potential to Obtain Higher Education" (in Russian). In Higher School of Economics, "Collection of Articles of HSE Graduate Students," 34–64. Higher School of Economics, Moscow.

Federal Institute of Pedagogical Measurement. 2009. "Results of Unified State Examination (May–June, 2009)" (in Russian). Federal Institute of Pedagogical Measurement, Moscow. p. 69. http://www.fipi.ru/view/sections/138/docs/478.html (accessed July 2013).

Kovaleva, G., Baranova, V., Koshelenko, N., Krasnianskaya, K., Krasnokutskaya, Krasnovsky, E., & Loginova, O. 2012. *Main Results of the International PISA Assessment 2009; Analytic Report* (in Russian). Moscow: MAX-PRESS.

Minister of Education and Science of the Krasnoyarsk Krai (Territory). 2010. "Implementing the National Education Initiative—Our New School: Prospects for the 2010–2011 Academic Year" (in Russian). Report presented at the August 2010 Teachers' Council, Krasnoyarsk. Krasnoyarsk Regional Ministry of Education and Science.

Municipal Institute of General Education of the Chuvash Republic. N.d. "Schedule for Preparing for the USE in the 2011–2012 School Year/ MOU Bichurga-Baishevskaya Secondary School" (in Russian). Municipal Institute of General Education, Cheboskary, Chuvash Republic. http://www.21419s02.edusite.ru/p213aa1.html (accessed July 2013).

Poldin, O.V. 2011. "Predicting Academic Performance in College Based on USE Results" (in Russian). *Applied Econometrics* 1 (21): 56–69.

Public Opinion Foundation. 2011. "Unified State Examination." *Dominanty* 23 (September 6, 2011). Public Opinion Foundation, Moscow. http://bd.fom.ru/pdf/d23ege11.pdf.

Tyumeneva, Y., and Y. Kuzmina. 2012. "The Effect of One Extra Year of Schooling on PISA Results: A Case of Countries with Different Tracking Systems" (December 21, 2012). Higher School of Economics Research Paper WP BRP 08/EDU/2012. Available on the Social Science Research Network (SSRN). http://ssrn.com/abstract=2192462 or http://dx.doi.org/10.2139/ssrn.2192462.

Tyumeneva, Y., and A. Valdman. 2013. "The First Data from the Comparative Analysis of the Results of the TIMSS 2011 and PISA 2012 Tests, Administered to the Same Sample of Russian Students." Paper presented at the international conference, "Russian Education in the Mirror of the International Comparative Studies," June 19–20, 2013, Russian Academy of Education, Moscow. http://www.centeroko.ru/conf2013/conf2013_eng.htm.

Useful websites

Center for Evaluating the Quality of Education (in Russian): www.centeroko.ru

International Association for the Evaluation of Educational Achievement: www.iea.nl

Organization for Economic Co-operation and Development Programme for International Student Assessment: www.oecd.org/pisa/

TIMSS and PIRLS International Study Center: timss.bc.edu

List of papers published in the current Working Paper series

1. Clarke, M. 2012. "What Matters Most for Student Assessment Systems: A Framework Paper." *(Available in Russian and Portuguese).*
2. Ramirez, M. 2012. "Developing the Enabling Context for Student Assessment in Chile." *(Available in Russian).*
3. Ramirez, M. 2012. "Disseminating and Using Student Assessment Information in Chile." *(Available in Russian).*
4. Liberman, J., and M. Clarke. 2012. "Review of World Bank Support for Student Assessment Activities in Client Countries." *(Available in Russian).*
5. Flockton, L. 2012. "The Development of the Student Assessment System in New Zealand." *(Available in Russian).*
6. Allen, R. 2012. "Developing the Enabling Context for School-Based Assessment in Queensland, Australia." *(Available in Russian).*
7. Castro, M. 2012. "Developing the Enabling Context for Student Assessment in Brazil." *(Available in Russian).*
8. Kanjee, A., and S. Acana. 2013. "Developing the Enabling Context for Student Assessment in Uganda." *(Available in Russian).*
9. Bolotov, V., G. Kovaleva, M. Pinskaya, and I. Valdman. 2013. "Developing the Enabling Context for Student Assessment in the Russian Federation."
10. Tyumeneva, Y. 2013. "Disseminating and Using Student Assessment Information in Russia."

Forthcoming papers

11. Senghor, Ousmane. "Disseminating and Using Student Assessment Information in The Gambia."

This case study examines how the Russian Federation disseminates and uses information from its student assessment system, drawing lessons for other countries seeking to more effectively use their own assessment data. Russia's Unified State Examination (USE) is primarily used for student selection and certification purposes, but a variety of other uses have been attached to it, including informing pedagogy, ensuring accountability, and monitoring education quality. This variety of uses has had both positive and negative consequences for the school system. Information from international large-scale assessments (e.g., PISA, TIMSS, and PIRLS) has been widely used to introduce reforms in the school system. Factors that have affected the differential use of USE and international large-scale assessment data include, among others, the purpose of a given assessment, its design features and level of credibility, and access to the assessment database.

Yulia Tyumeneva, *Head, Educational and Psychological Measurement MA program, Higher School of Economics, Russia*

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

The Russia Education Aid for Development Trust Fund is a collaboration between the Russian Federation and the World Bank that supports the improvement of student learning outcomes in low-income countries through the development of robust student assessment systems. Visit the READ website at www.worldbank.org/readtf for additional information.