

# The Past, Present, and Possible Future of the Russian Education Assessment System

**[Viktor Bolotov](#)**

Received in  
March 2018

---

**Viktor Bolotov**

Doctor of Sciences in Pedagogy, Professor, President of the Eurasian Association for Educational Assessment. Address: 21 Novy Arbat St, 119019 Moscow, Russian Federation. Email: [vikbolotov@yandex.ru](mailto:vikbolotov@yandex.ru)

**Abstract.** The article describes the key stages in the development of the educational assessment system in Russia: certification of educational institutions, participation in international comparative studies, implementation of the Unified State Examination (USE) and Basic State Examination (BSE), and emergence of a community of education assessment experts. The most urgent goals in the development of the Russian education assessment system are seen in switching to competency-oriented USE and BSE (with the

subject-specific component preserved), developing national monitoring studies to compare education quality across regions and municipalities, tracing the socialization patterns of school graduates, elaborating various models of in-class and in-school assessment, and providing tools to measure individual progress of students. Meanwhile, the lack of competent interpretation of measurement results appears to be the main challenge in educational assessment.

**Keywords:** schooling, Russian educational assessment system, certification of educational institutions, international comparative studies, USE, BSE, monitoring, ratings, rankings, Russian Nationwide Tests.

**DOI:** 10.17323/1814-9545-2018-3-287-297

---

Education quality has been receiving more and more attention recently, a transition being made from supervision and control to education quality management. Naturally, more and more speculations arise around the issue, so I find it desirable to look into the history of its evolution. School education will be my sole focus, as preschool, vocational and tertiary education should each be described separately.

## **The Key Components of the Russian Education Assessment System**

Long before the very notion of the *Russian education assessment system* was adopted by educators, some components of this system had already been discussed. Back in the early 1990s, World Bank experts from the Netherlands and Great Britain proposed analyzing the existing national examination and monitoring systems and attempting to create one in Russia. That was a pretty inconceivable scenario, given that teachers in Russia were hardly even paid their salaries at the time.

Nevertheless, the 1992 Law on Education already defined the idea of *certification of educational institutions*, which implied establishing how well the quality of graduates complied with the requirements stipulated in the national learning standards (this term was also defined for the first time in the law of 1992). An institution was considered certified if at least 50 percent of its graduates demonstrated the knowledge and skills required by the standard. It was not until 2004 that the first learning standard was actually developed (the reasons for this are beyond the scope of this article), and still institutions were certified on a regular basis. Every region came up with their own graduate assessment materials to test conformance with the basic education plan and program recommendations—instead of learning standards. The quality of those materials was non-negotiable, and there were no experts to design and evaluate measurement tools and procedures. Consequently, any general speculations on the quality of schooling were out of the question.

Russia became a regular participant in international comparative studies in 1995, beginning with the Trends in International Mathematics and Science Study (TIMSS). It has also participated in the Programme for International Student Assessment (PISA) since 2000 and the Progress in International Reading Literacy Study (PIRLS) since 2001. In addition, there were now experts who could create and test measurement tools and procedures. Russia performed great in the TIMSS and PIRLS, scoring below the OECD average in the PISA. Results obtained by Russian students in the international assessments were used to develop recommendations for curriculum methodologists, upgrade advanced teacher training programs, and adjust the content of textbooks, e. g. by adding tasks testing various literacies and competencies. Later on, the approaches adopted for the development of international assessment tests were used to design the standards of school education. It can be said therefore that the results of international comparative studies became a tool of school education quality management in Russia.

In 1996, centralized school-leaving testing was introduced, which used unified measurement materials (multiple-choice tests) and a unified procedure to exempt school graduates from having their knowledge of school subjects tested twice within a month (first when leaving high school and then when entering college). Since centralized testing was not obligatory for school leavers (moreover, participation was fee-based) while schools and colleges were the ones to decide whether to credit the scores as passing for graduation or admission, there was not much debate over the quality of test measures or objectivity. Colleges mostly accepted centralized testing scores in non-major courses, e. g. a lot of engineering universities happily credited scores in Russian language. No inference on education quality could possibly be drawn from the outcomes of centralized testing.

*Translated from  
Russian by  
I. Zhuchkova.*

It was only in 2001, when the experiment with the Unified State Examination (USE) began to unfold, that education quality started becoming a big deal.

It might be useful to remind the reader of what actually prompted the emergence of the USE. By the year 2000, it had become blatantly obvious that grades in the certificates of graduates from even two neighboring schools did not make it possible to compare the level of those graduates' knowledge and competencies. In a situation where the number of gold medalists was forced up to boost privileges in college admission, it was no use talking about the quality of school education.

College admission procedures were no less a matter of concern. Every college designed entrance examinations of their own, which led to proliferation of 'under-the-counter' college prep tutoring, meaning that college tutors and prep courses only taught the topics that the candidates would find in a specific exam. There was a huge mass media coverage of abuse in entrance exam procedures, when pre-tutored candidates were guaranteed admission and everybody else could only make it into the college if there were still vacancies left. Many colleges also had secret understandings with specific schools, accepting results of their low-supervised exit examinations to admit graduates. Clearly, a child 'off the street' had almost no chance of being enrolled in a school like that. As a result, intra-national student migration rates plummeted in the top universities of Russia, where only 25 percent of newly-enrolled students were non-residents in 2000, as compared to the Soviet rate of 75 percent. The proportion of rural students in regional colleges reduced a lot, too.

The USE was designed to provide assessment of individual attainment of high school graduates and thus college applicants in school subjects regardless of which school they studied at or which college they applied to. A study of global practices followed by extended discussion resulted in a decision that two school exit exams, Russian and mathematics, will only be available for taking as USE tests, while students could still choose between the USE and conventional examination format for the rest of the subjects. Understandably, their choice was predetermined by the array of admission tests in every specific major in every specific college. As soon it was all set with the mandatory and optional exams, it was time to decide on the format of questions. Again, analysis of the results of a number of international assessments as well as national tests and monitoring studies in different countries, first of all Australia, Great Britain, the Netherlands and the United States, resulted in the following structure of test materials: Part A contained multiple-choice tests, Part B contained short-answer questions, and Part C suggested giving extended answers (argumentative, essay, problem solving, etc.). Part A and B tasks were checked by a computer, whereas the checking of Part C tasks was carried out by experts.

While international experience could be used to develop the structure of test materials, the policies and procedures were created entirely from scratch: there had been no precedent of virtually simultaneous countrywide testing in a country covering ten time zones.

When the USE integration experiment was launched in 2001, only four federal subjects of Russia took up voluntary participation in it, test materials were available for eight school subjects only, the scores were accepted by 16 colleges, and 45,000 man-exams were conducted. The aim of the experiment was to hone the technology used for designing the test materials and procedures, from the unified regulations for interaction among specifically-established federal and regional agencies, through involvement of public monitoring groups, to the appeal investigation procedure. In 2008, when the experiment stage was nearing its end, the Unified State Examination involved 84 federal subjects of Russia as voluntary participants, featured test materials in 13 school subjects, had its results accepted by 1,800 colleges and their branches, and boasted a record of 2,665,000 man-exams successfully conducted.

Alongside the USE, the middle-school student academic achievement test has been in place since 2003 (originally called the State Final Examination and then renamed into the Basic State Examination). Test materials were developed at the federal level, but, unlike with the USE, the exam procedure was designed and supervised by the relevant federal subjects—as interregional mobility is extremely low among middle school graduates, there is practically no concern about the equivalence of assessment results across the federal subjects.

Thus, the key component of the Russian educational assessment system was constructed, that being the independent assessment of academic achievement of middle and high school graduates. The process of construction involved establishing a network of regional information processing centers, which later evolved into centers for educational assessment. In addition to making quality higher education more accessible to children from remote regions and rural localities, the USE performed another important function of providing teachers and curriculum methodologists with valuable information on the level of understanding of specific topics within the subjects included in the USE. Annual reports contained test materials and detailed analysis of student performance, which were then actively used for the purposes of advanced teacher training. Similar activities were carried out at the levels of federal subjects, municipalities and individual schools, which undoubtedly had an effect on the quality of Russian education.

A fact that cannot be ignored though is that USE and BSE results have been continuously misused since the first year of the experiment despite federal education authorities issuing a number of messages, explaining that such practices are unacceptable. Results of high-stakes testing—which is a term adopted for tests with important consequences for the test taker (in this case, in terms of obtain-

ing a school leaving qualification and/or entering college)—began to be used to directly assess the performance of teachers, schools, municipalities, regions and even governors as well as to build ratings of all sorts. However, evaluating performances of any type without making allowance for the context (socioeconomic status, whether home language is different from the language of instruction, the level of educational infrastructure available, etc.) is simply wrong and fraught with punishment of the innocent and reward for the uninvolved. In fact, analysis of test performance shows that nearly all 100-point scorers use out-of-school resources to prepare for the test (dedicated courses, including those online, tutorship, etc.) and students scoring below the required minimum often come from low socioeconomic backgrounds and live in depressed or remote districts. Misuse of USE and BSE results prompted schools, municipalities and regions to ensure high student scores at all costs, which resulted in numerous attempts to falsify test results.

Since USE-based evaluation of governors and later mayors was abolished (regions' average USE scores used to be a criterion of governor and mayor performance) and as a result of the unprecedented measures taken over the last two years to enhance security during examinations, objectivity and reliability of the test results have improved dramatically.

It has to be admitted, however, that misuse of USE results has some substance behind it: decision-makers simply do not dispose of any other remotely reliable information that would allow them to assess the effectiveness of teachers, schools, municipalities, etc. For this reason, the concept of the Russian Nationwide System of Educational Assessment was developed in 2006, yet it was never approved due to staff changes in the industry. Along with improving the USE and BSE measures and procedures (including discrimination between the basic and advanced levels in obligatory subjects), the concept implied creating national studies of not only academic achievement but also socialization of students and school leavers, testing diverse models of in-school assessment, and designing programs to instruct teachers and school administrators to use and interpret results of different types of tests with a view to improving education quality. A number of the concept provisions became part of the 2012 Law on Education and other regulatory documents, which certainly fostered the evolution of the Russian educational assessment system.

An important role in this process was played by Russia Education Aid for Development (READ), a project launched in 2008 as part of cooperation between Russia's government and the World Bank to support developing countries. One of the paramount goals of the project was to foster a professional community that would deal with education quality problems and carry out tests, assessments and studies in Russia and the CIS countries. Measures used to achieve that goal included conducting regular webinars on the pressing issues of educational

assessment, holding dedicated conferences, and publishing a journal. Events held as part of the project were attended by hundreds of professionals from all over Russia and the CIS countries. READ assisted the development of a number of measurement materials that can be used to evaluate and monitor both the individual progress of students and the performance of educational institutions. Unfortunately, unlike their regional-level colleagues, federal education authorities showed very little interest in the project.

**Monitoring in Educational Assessment**

Most researchers understand monitoring as regular observations and description of the state of an object(s) using a small number of specific parameters. Monitoring results are often presented as ratings (rating being understood as one-dimensional comparison by a pre-selected criterion).

Educational ratings are normally used to identify and spread the best practices as well as to spot risk zones, i. e. schools, municipalities or federal subjects with consistently low learning outcomes, which require targeted action plans to improve the situation.

In Russia, educational monitoring is represented first of all by various international assessments, of which Russia has been a regular participant and the results of which have been used to develop recommendations on improving the quality of Russian education. The integration of the USE was immediately followed by attempts to use its outcomes as the basis for monitoring the quality of school education at the national, regional and municipal levels, which resulted in a shower of ratings of all sorts. Two fundamental errors were committed along the way. First, everyone ignored the fact that high-stakes tests cannot be used as indicators of schools', municipalities' or regions' performance unless contextual parameters are taken into account. Second, no allowance was made for the fact that USE scores of different years just cannot be compared due to the changes in test design (the division into the basic and advanced levels in mathematics and abolishment of multiple-choice tests for no good reason) and to the test conditions growing ever stricter over the recent years (which is totally justifiable). Otherwise speaking, one cannot judge on education quality improving or worsening from a mere comparison of USE performance in different years. What is more, the "politically" motivated refusal to use multiple-choice tests, which have been and will continue to be widely used in all international studies, has resulted in the minimum passing USE scores permanently going down.

Ratings based on results from the same year yielded no more information as they did not differentiate between schools, municipalities and regions with different socioeconomic and cultural parameters. The resulting ratings were topped by the so-called "governor's schools", while schools educating predominantly students from low socioeconomic backgrounds were lagging far behind. It would hardly

make sense to talk about spreading the best practices of the “governor’s schools” to underfilled rural schools in depressed districts as an implication of such ratings. Consequently, it would also be very hard to contemplate any education quality improvements based on them.

It was only in 2014 that the National Survey of Education Quality (NSEQ) was launched, implying regular assessments of the quality of schooling in specific subjects, at specific levels of school education, in specific classes. However, the implications of the findings remain utterly limited. Results of the NSEQ, performed at the national level, cannot be compared to any global outcomes. Localized school samples are not representative of the federal subjects of Russia, making it impossible to compare performance or draw generalized conclusions on the quality of education in regions, municipalities and individual schools. Besides, no fixed intervals for NSEQ in particular school subjects have been established so far, which makes it impossible to monitor trends in subject-specific teaching. As a result, it is not so clear who can use NSEQ findings to improve education quality and how.

A number of regions conduct monitoring assessments of their own, which most often use the READ tools or those applied by the Center for Educational Assessment of the Institute for Strategy of Education Development.

### **Current Issues in the Development of the Russian Educational Assessment System**

I will start with the innovation that has been probably the major concern in the educational community lately. The Russian Nationwide Tests, which represent tests for school students at the end of every year of schooling, have been the most massive evaluation procedure in the Russian education system. About three million school students and nearly 40,000 schools took part in the first round in April–May 2017, which only involved 4th, 5th and 10th grades. The way this assessment procedure is organized raises a whole lot of red flags.

First, the Russian Nationwide Tests are claimed to use the tasks developed at the federal level in compliance with the Federal State Education Standard and to provide uniformity of assessment criteria, although schools are the ones in charge of the tests. However, the standards are based on stages, not grades, and do not contain subject-specific performance requirements, so the law allows every school to develop education programs of its own. That means, in fact, that the Russian Nationwide Tests will again overregulate school activities and experimental schools will again be stigmatized as non-compliant. As for unified criteria, the experiences of international Baccalaureate schools and of USE integration have made it obvious that a great deal of effort should be invested in elaborating such criteria and teaching in order to use them.

Second, the implications of the Russian Nationwide Tests’ results remain unclear. Allegedly, teachers are supposed to use them to evaluate academic achievement of students and develop individual learn-

ing plans. In other words, the tests provide material for diagnostics, thus being a service for teachers and schools. But why then should the procedure be regulated that much? This degree of regulation may result, and it already does, in teachers and schools being compared on the basis of performance in the Russian Nationwide Tests. Fraud and punishment of the innocent are predictable consequences, as the tests are run by schools. It is also very likely that the results will be used for rating.

That way, the Russian Nationwide Tests turn out to be a weird hybrid of materials for in-school assessment (but why then so much external regulation?), monitoring (but why then test every single child? And how could it be viable without external control?) and high-stakes testing (because there will inevitably be high and low achievers).

This assessment monster consumes an awful amount of finance, time and human resources, diverting attention from the actually challenging issues in the Russian educational assessment system, which I believe include the following:

- Transition to competency-oriented USE and BSE, while keeping in line with the middle and high school standards and preserving the subject-specific component;
- Development of national monitoring studies to compare education quality across regions and municipalities, not only by the level of students' knowledge but also by the level of soft skills they have developed, and monitor the socialization patterns of school graduates and at-risk teenagers;
- Elaboration of various models of in-class and in-school assessment and creation of tools to measure individual progress of students, not only in subject-specific knowledge but also in terms of how they develop various competencies.

Meanwhile, the lack of competent interpretation of measurement results at all levels appears to be the main challenge in educational assessment today. The critical step in using test findings consists in switching from ratings to rankings, i. e. multi-parameter comparative assessments that allow users to sort the assessment results by any parameter that might be of interest to them and thus provide for a whole array of ratings, and preparing thoughtful managerial decisions on education quality enhancement based on such rankings.

Some really good practices have been developed in all of the abovementioned fields, but they are all results of enthusiastic effort and never receive the priority attention from education authorities. Therefore, the prospects for the development of the Russian educational assessment system depend on solving the problems described above.

As for the Unified State Examination as the central component of this system today, I speculate that it will disappear in its current



form, surrendering much of its infrastructure to centers that will carry out independent evaluation of the level of subject-specific knowledge, various literacies and soft skills (the OECD and the WorldSkills have already started working in that direction). Such centers will issue certificates that people will collect into portfolios and use when enrolling in postsecondary education and competing for jobs. Some prototypes of such centers already exist, one of those being TOEFL testing locations.

Overall, the weight of education quality assessment issues is certain to continue growing in the education system evolution agendas, both in Russia and globally.