Using Contextual Data in the Education Quality Assessment System:

Tools Development and Testing Practices

G. Yastrebov, M. Pinskaya, S. Kosaretsky

Gordey Yastrebov
Candidate of Sciences in Sociology, Senior Research Fellow, Laboratory for Comparative Analysis of Post-Socialist Development, National Research University—Higher School of Economics; Doctoral Student, European University Institute (Florence, Italy). E-mail: g.yastrebov@hse.ru

Marina Pinskaya
Candidate of Sciences in Pedagogy, Leading Research Fellow, Center of Social and Economic School Development, Institute of Education, National Research University—Higher School of Economics. E-mail: m-pinskaya@yandex.ru

Sergey Kosaretsky
Candidate of Sciences in Psychology, Director, Center of Social and Economic School Development, Institute of Education, National Research University—Higher School of Economics, Russian Federation. E-mail: skosaretski@hse.ru

Address: 20 Myasnitskaya str., Moscow, 101000, Russian Federation

Abstract. The paper proposes an approach to assessing the performance of educational institutions with regard to their specific social features. To develop this approach, the authors relied upon 1) results of numerous studies proving correlations between student performance and contextual factors (both in Russia and abroad); 2) foreign colleagues' experience of solving similar problems; and 3) the idea of providing minimum required information to enable such assessments in contemporary Russia. The fundamental idea behind the proposed assessment tool is that, having the necessary data at hand, one can empirically identify stable correlations between student performance and contextual factors (e. g., the social composition of students). In research practice, these correlations were revealed through multiple regression analysis. The results of such analysis—established empirical correlations—may then be used to "discount" formal progress; that is, to have justifiably higher expectations about institutions in more favorable contexts and lower expectations about those in less favorable situations. The authors consider two ways of using this information—one based on a formula and one based on a specific index (the index of school social well-being) that they have developed. They also draw attention towards possible constraints associated with using these tools and touch upon the more global problem of considering contextual factors in assessing the quality of education in Russia.

Keywords: education quality, quality assessment, contextualization, contextual data, social context, social composition of students, the Unified State Examination (USE).

There is an established approach to assessing the performance of educational institutions that involves analyzing their working conditions, or the school context. The problem of transforming this approach remains unsolved, despite its seemingly simple and obvious nature. Several years of research based on data provided by Russian schools proved that the correlation between academic achievement and external, independent factors shaping the conditions for school activities, which is the focus of education policies and management strategies all over the world, is also a crucial issue for Russian schools. The characteristics of school resources and school population play an important role in forming the school context. Although the results of such research have been widely presented on various social platforms[1] and debated in professional discussions[2], no firm conclusions have yet been made in this regard.

Among the documents regulating the assessment of the performance of educational institutions, the most noteworthy is the Action Plan ("workflow chart") "Changes in Social Industries Designed to Enhance Performance of Education and Science" (Order No 2620-p of the Government of the Russian Federation dated December 30, 2012), which envisages the "development (modification) of indexes to monitor performance of subordinate state (municipal) general education institutions, their principals and employees of major categories." The notion of "performance" was offered for the social sphere and for science as a whole, but its scope and specificities as applied to schools were never studied in depth or brought up for serious discussion. As a result, the term unfortunately lost its pedagogical component. As required by the Plan, these indexes were hastily developed in 2013 by local governments and state governmental authorities that were in charge of education in federal subjects of Russia. Therefore, the indexes represent rather eclectic kits with only rough assessment procedures and usage to encourage directors and employees.

Decree No 597 of the President of the Russian Federation ("On Measures to Implement the National Social Policy"), dated May 7, 2012 was issued to ensure the "development of an independent system to assess performance of organizations providing social services, which includes defining of performance criteria and introduction of public rankings for such organizations" (p. 1 "к") by April 1, 2013.

This requirement entailed an uncoordinated search for performance assessment criteria, again without going deep into the notion

---

[1] For example, public hearings *Equalizing Children's Opportunities for Quality Education* held by the Commission on the Development of Education of the Civic Chamber of the Russian Federation (June 2012) [Frumin et al., 2012].

[2] In particular, the special 9th International Research and Practice Conference *Trends in Education Development: Equal, Even Though Different. How to Close the Gap in Access to Quality Education*, February 17–18, 2012.

itself. Public rankings appeared to be more recognizable and "convenient," which made them the focus of the developers of the quality assessment system. In fact, rankings were substituted for analysis and well-directed management strategy. Rankings allowed policymakers to proceed with the practice widely used since the implementation of the national project "Education," one that identified and supported the "leaders"—the most powerful and successful educational institutions that are usually rich in resources—while ignoring the low-ranking organizations that demonstrably need a high-quality assessment of their situations in order to secure sources of support. The simplest "linear" ranking models based mostly on scores obtained in the USE and in academic competitions makes it impossible to assess the real efforts made by schools. As a result, some truly efficient educational institutions that achieved the highest results possible given the contexts in which they work are left neglected.

Another document, Resolution of the Government of the Russian Federation No 662 "On Monitoring of the Education System" dated August 5, 2013, stressed the importance of "continuous and consistent analysis and evaluation of the current state and prospects of education (also in terms of performance of educational institutions), as well as enhancement of the productivity of the education system through the improvement of the quality of related managerial decisions." The list of information about the education system subject to obligatory monitoring includes some school resource data but omits the most essential characteristics of student population. Such an approach sets back severely the "improvement of quality of managerial decisions,"[3] as stated in the Resolution.

We are forced to admit that present-day practices of assessing school performance and using the results in educational management still:

- do not take into account the differences in school resources and in characteristics of student population;
- focus on the top-ranking institutions, while outsiders and low-rankers are disregarded by the education authorities;
- do not evaluate the contribution and effort made by schools;
- measure the construct of "performance" using guesswork.

However, the scope of the term "performance" with regard to educational institutions and assessment of the conditions (context) of their operation remains of vital importance to federal, regional and local education authorities. The national program "Development of Education" for 2013–2020 includes Subprogram 3 ("Enhancing the assess-

---

[3] The Rules of Education System Monitoring, par. 2 http://www.rg.ru/2013/08/19/monitoring-site-dok.html

ment of performance and information transparency of the education system"), which implies the introduction of a national education quality assessment and monitoring research system. The Ministry of Education and Science of the Russian Federation together with the Federal Agency of Supervision in Education and Science (Rosobrnadzor) continue to work on the development[4] of a national system of general education quality assessment.

Furthermore, a number of subjects of the Russian Federation have expressed interest in going beyond simply following the "workflow chart" by the book and instead creating a truly reasonable school performance assessment system.

In this paper, we offer an approach to considering contextual data in educational organization performance assessment that may be of use in reaching the abovementioned goal.

1. Considering Contextual Data in Education

In stating that there are no established practices of considering contextual data in the education quality management system and in the assessment of educational institution performance, we do not mean that contextual information is completely and ubiquitously ignored when educational institutions or systems of education are compared. Nevertheless, the existing studies and attempts are 1) inconsistent 2) usually cover only some of the regions, and 3) the methods used are so diverse that meaningful comparison becomes impossible (see, for example, [Sobkin, Pisarsky, 1998; Konstantinovsky, Vakhshtayn, Kurakin, 2013; Yastrebov et al., 2013]).

Still, what do we mean when we talk about contextual data in education? What is context? In the broadest sense, context is an environment or a system of circumstances that form the setting for a process or a phenomenon, giving them a particular twist. In everyday life, when we say that something is "torn out of the context," we mean that we are unable to understand the information fully until we have additional data that will make this information meaningful from the point of view in which we are interested.

In terms of education, context may be defined as circumstances that form the setting for the educational process but which are external to the process; that is, they are not integral to the process but still affect it to a large extent. In fact, context is a set of external but often unobvious factors that players of the educational process cannot manage but should not ignore when assessing the results of the process.

In our earlier publications, we dwelled on the theoretical grounds of considering contextual data in education [Pinskaya, Kosaretsky,

---

[4] A project of the OSOKO concept (the All-Russian System of Education Quality Assessment) was drafted in 2007 but has not yet been implemented at the national level.

Frumin, 2011; Yastrebov, 2012; Yastrebov et al., 2013]. It will be sufficient to say here that most contemporary studies on education organization in the context of social and economic inequality inevitably refer to the leading sociologists, such as John Coleman, Pierre Bourdieu, and their numerous successors. They have proved decisively in their studies that socioeconomic factors that characterize students' families (or, on a larger scale, local communities, districts, or territories) determine the context which is absolutely indispensable to describe the operation of individual educational institutions or of the entire education system and, consequently, to evaluate their performance correctly.

Ample research has been conducted to find out *how* socioeconomic advantage is translated into educational advantage. For instance, the well-off can afford exam coaches and other forms of supplementary education. Highly educated families contribute to the educational advantage of their children by providing a specific culture at home, as well as by being more concerned with efficient interaction and active engagement with their children's schools. Strange as it may appear, despite the abundance of empirical studies devoted to discrimination in Russian education [Konstantinovsky et al., 2011; Yastrebov, 2010; Prakhov, Yudkevich, 2012; Roshchina, 2012] and a rather clear idea of how this discrimination is reproduced in Russian management practices, this knowledge is little sought for, almost unused, and, most astoundingly, still discarded by statistical accounting.

Meanwhile, the practice has been successfully applied in a number of countries. For example, the UK's school performance assessment system uses data on age, gender, ethnicity, and the socioeconomic status of students while also taking into account possible mobility between educational institutions that may cause distortions in the assessment[5]. The United States do not have a unified system of considering contextual data, but some states (e. g., Florida, South Carolina, Wisconsin, and Tennessee) are promoting their own education quality assessment programs based on similar principles [OECD, 2008. P. 76]. Readers of the *Problemy sovremennogo obrazovaniya* magazine have probably heard a lot of the Australian [Valdman, 2013a] and Chilean [Valdman, 2013b] experience. Australia provides an integrated consideration of contextual information with the help of a specifically designed index (Index of Community Socio-Economic Advantage), which is calculated for each school individually. The index sums up socioeconomic indicators[6] whose proportions in the

---

[5] All information about schools and assessment procedures is posted publicly on the website of the UK Department for Education: http://www.education.gov.uk/schools/performance/

[6] In particular, the index includes data on parent employment and education, intrafamilial language, belonging to indigenous Australians, type of community where the school is located.

formula are annually calculated using special analysis to show the extent to which this or that indicator is responsible for variations in student achievement. Next, educational institutions are compared according to their performance using the index under the statistical neighbor model, i. e., being grouped based on similar index rates (and, hence, similar socioeconomic situations). In Chile, all schools are divided into five groups using cluster analysis with three variables: parent education, household income, and social vulnerability index[7]; further comparison of educational organizations is also performed within the sorted groups[8].

Sets of applied variables are essentially the same in most countries and, apparently, are easy to implement in statistical accounting, as required information is often collected by schools themselves at the admission stage. Discrepancies in accounting systems may be related to the importance attached to specific variables in different national contexts. For example, recording the ethnicity of students and migration status of their families may be irrelevant in countries that are not facing massive influxes of immigrants. Financial, legal, moral and ethical restrictions on obtaining some of the contextual information necessary for the purposes of education quality assessment are also of great significance. These constraints are gradually being overcome to ensure efficient education quality management.

The necessity of considering contextual variables when comparing performance of educational institutions and education systems has been addressed recently by renowned Russian experts [Agranovich, 2008; Bochenkov, Valdman, 2013; Bolotov, Valdman, 2013]. The authors of the articles included in the collection "Variable- and Index-Based Management of Education Quality in Regions" compare regional education systems and call attention to the following contextual variables that determine the socioeconomic features of the federal subjects of Russia: per capita gross regional product, proportion of rural population, migration balance, unemployment rate, indexes of sociocultural infrastructure development, etc. [Agranovich et al., 2008. P. 21]. A similar philosophy underlies one of the pioneering attempts to systematize regional educational situations in Russia [Sobkin, Pisarsky, 1998].

There are few existing examples of using contextual information directly to create a typology of educational institutions. One such example is a study headed by D. Konstantinovsky, which uses cluster analysis to develop a real typology. This approach enabled a classification that considered several dimensions: apart from "output" var-

---

[7] The poverty index from official statistics that is calculated to determine the size of school lunch financing.

[8] For more information on how contextual data is used in education statistics by other countries, see one of our earlier publications [Yastrebov et al., 2013. P. 194–195] or the relevant OECD report [OECD, 2008. P. 135–137].

iables, those of "input" and "process" were also analyzed. The authors undoubtedly did an innovative thing by selecting variables not only based on theoretical models but also using data obtained in statistical analysis which aimed at defining the contextual characteristics that correlated most strongly with academic performance [Konstantinovsky, Vakhshtayn, Kurakin, 2013]. Other authors have also proposed using the cluster approach to build a typology of educational institutions that takes contextual information into consideration [Agranovich et al., 2008].

Clusterization is certainly a step in the right direction. Whatever the method, statistical (cluster analysis) or random (with predetermined typological criteria), the most essential is the idea that at the "output," we have groups of objects that have similar *sets* of characteristics. However, we believe the approach has a meaningful constraint in the discreteness of the clusters and types identified. By definition, clusterization reduces the spectrum of objects to a limited number of groups, thus inevitably omitting some of the information about their characteristics. This is true primarily for characteristics measured with continuous variables, such as USE points, size of the school population, and the proportion of students from low-income families. Differences between objects within a type are practically ignored, making more diversified assessment and comparison rather difficult (especially for objects with boundary values within their groups).

In this paper, we propose a method of using contextual data to compare performance of educational institutions that circumvents the restrictions mentioned. This method attempts to adapt the existing methodological considerations in this field to the Russian environment.

2. Conceptual Approaches to School Performance Assessment

Ample research on the factors determining academic achievements and significance attributed to enhancing the performance of education systems all over the world has provided a conceptual foundation that can be used to assess and compare educational institutions.

Many conceptual models have been developed to describe the process of education and the factors potentially affecting its efficiency [Rumberger, Thomas, 2000; Shavelson et al., 1987; Willms, 1992; Barr, Dreeben, 1983]. All of them regard this process as a multi-factorial phenomenon comprising a number of mutually dependent circumstances.

The multilayer structure of the learning process absolutely needs to be considered, as learning outcomes are produced by factors of different levels layered one on top of another. Thus, academic achievements at the personal level depend on the efforts made by the student and the amount of practice they receive doing a certain amount of homework. At the classroom level, achievements also de-

pend on the amount and quality of teaching and the teacher's ability to provide a good learning environment. Finally, at the school level, important roles are played by available financial, staff and other resources, as well as by the type of organization, factors which also account for the amount and quality of education opportunities available to students. Therefore, using student achievements per se as a performance criterion, without having a clear idea of which of the numerous factors are responsible for it and at which levels, is not a reliable approach, not to mention the fact that the practice of constructing multiple rankings ignores the context in which the achievements were made.

Some of the foreign studies on school performance (e.g. [Hanushek, 1986; Levin, 1994]) are centered around the so-called *business model of school* that treats educational institutions as any other organizations conducting useful business in the context of a specific set of predetermined constraints. This model sees the process of education as consisting of three elements.

First, *resources*, which are not restricted to teaching staff and material resources but also include the student body, which is crucial. The latter is characterized by the proportion of talented students, the level of their initial competencies, and other school-independent variables affecting knowledge of the curriculum (for example, willingness of parents to pay for supplementary courses and coaching). Thus, resources are treated as predetermined, or external to educational institutions and the conditions of their operation [Hanushek, 1989].

The second element is the very *process* of education, in its broadest sense. The nature of this process is what determines how efficiently the resources available are translated into academic achievements. The quality of the process may be related to the effectiveness of school management, specific features of the school climate, use of certain teaching practices, or the relevance of learning programs. In fact, describing these factors is the key to identifying the productive learning process models [Shavelson et al., 1987].

The third and final element is *outcomes*, i.e., the level of knowledge of the curriculum, academic performance, and other indicators of student achievement. The selection of variables comprising education outcomes depends on the a priori idea of the performance function of education or of the function of a specific educational institution (e.g., socialization, mastering important social competencies, etc.). To identify truly productive learning process models, it is essential not to confine oneself to outcomes but also to consider the reference conditions in which they were attained.

The necessity of assessing school performance led to the development of special tools—a unique class of statistical models allowing for a direct or indirect quantitative evaluation of parameters that characterize contributions in academic achievement or education outcomes made by different factors. This class of models feeds

upon the principle of statistical description of one (response) variable through others (explanatory) using the *multiple regression analysis* methods. These methods are among the standard competencies that are part of most modern higher professional education curricula both in natural sciences and some of the social sciences (primarily economics and sociology).

Specific models, or tools, are conventionally broken up into three main groups according to the availability of variables used to measure academic achievements or educational outcomes:

1) current performance assessment models;
2) academic progress assessment models (value-added models);
3) models assessing academic performance measured with discrete variables [Rumberger, Palardy, 2004. P. 240].

The choice of the model always depends on the nature of the data available to researchers or authorities performing the assessment.

School performance is most often measured with *current performance assessment models*. This type of model is used when information about academic achievements is only relevant for a specific point in time and there is no way to evaluate the dynamics (for instance, by retesting or different tests after a certain period of time). This can be illustrated by a model analyzing scores achieved by a student on the Unified State Exam (USE), the State Final Examination (SFE), or any other test as a function of their performance in the classroom, commitment to homework, quality of teaching, characteristics of the curriculum, etc. At the level of schools (not individual students), current performance assessment models can be ones that assess the correlation between scores of a specific student cohort with parameters of its staffing and financial support and the social composition of the student population (models of contextualization).

Academic progress assessment models (value-added models) are currently less widespread, but they are among the most advanced models used to assess performance of educational institutions [OECD, 2008]. They are used quite rarely due to the sophisticated and costly methods of obtaining the data required for their construction—the assessment of academic progress requires continuous testing that would make it possible to measure accumulation of knowledge over a student's period of study at school. Obviously, very few countries can afford such assessment systems. They are most widely used today in economically developed countries like the UK and some states in the U. S. The undisputable benefit of these models is that they provide the most precise assessment of progress achieved in schools because of the possibility of documenting students' initial levels of knowledge.

Finally, there are *models assessing academic performance measured with discrete variables,* which are used in rare cases and

in which the resulting variable is expressed as a discrete value. This may be, for instance, expulsion from school, voluntary transfer from one school to another, or winning in an academic competition. Another possible example of a discrete variable is the translation of education results to a limited ordinal scale such as the current 1 to 5 performance rating scale widely accepted in Russia. As applied to educational institutions themselves, a discrete variable may be, for example, an institution's position in a formal performance classification. Generally speaking, these models are applied when educational results for some reason cannot be measured using continuous variables, such as the USE or SFE points.

No matter which model is used to assess academic achievement, the critical objective is to identify and consider factors at different levels: *personal* (competencies, family characteristics), *group* (class composition, teachers' characteristics), *school* (overall staffing, financial and curricula situation) and, finally, *geographic* (characteristics of communities, districts and even regions where the schools are located). This approach is unique in that it evaluates the contribution of each factor (or set of factors) individually, analyzing them in a specific order, from elementary to the most aggregate level. In particular, this analysis lets us determine the level which is prevalent among the factors material to outcomes. The information is priceless for those in charge of education policies and efficient reallocation of resources. No wonder this multilevel design has been applied to most studies and analytical reports that develop the tradition of school performance analysis abroad [Aitkin, Longford, 1986; Willms, Raudenbush, 1989; OECD, 2008. P. 15].

At the same time, even simpler models using summarized information about educational institutions (aggregate characteristics of education outcomes, social composition of the student body, staffing and material resources of the school, etc.) can easily demonstrate how inadequate "output"-based school ranking is [Dyer, Linn, Patton, 1969; Burstein, 1980; McCall, Kingsbury, Olson, 2004]. The idea behind this analysis is intuitive: schools are ranked not by absolute amount or level of education outcomes, but by how much these results are higher or lower than the overall average values for schools *with identical resource and student body parameters*. This difference between observed and average results is often interpreted in literary sources as "school effects" or "school contribution." However, these definitions cannot be considered correct anymore, as there are now more precise models that take into account academic progress (value-added models) [OECD, 2008. P. 15].

3. Experience of Using Contextual Data to Compare Performance of Russian Schools

Not only does the system of statistical accounting applied in Russian schools not include individual indicators of academic progress, but it also does not provide for analysis of individual achievements in rela-

tion to socio-demographic data (although the latter may be used in full confidentiality for the purposes of such analysis).

Statistics about e-monitoring under the education initiative "Our New School"[9] are, perhaps, the most comprehensive and transparent ones today. However, educational institutions form the basic level of data interpretation, i. e., no student-specific information is available; more critically, the published spectrum of variables does not include social composition characteristics despite its diversity[10]. Thus, even with the opportunities provided by this database, it is impossible to perform an even slightly adequate analysis of school performance in accordance with the principles and approaches described above. The unavailability of required contextual information—notably social characteristics of students and their families—is a fundamental problem that must be solved before any school performance assessment models are adopted in Russia.

With the intent of demonstrating how contextual data can be used to compare performance of Russian schools, we will refer to our previous study in which we analyzed information collected from two regions of Russia in 2011–2012 to prove that academic performance demonstrably differs among schools with different student body social compositions and different staffing and material resources [Yastrebov et al., 2013]. We obtained the data for that study from the so-called "social passports" of educational institutions, which are not yet obligatory for schools to have and are only used in a limited number of subjects of the Russian Federation. In our case, the data was collected on request, after appealing to regional education authorities.

Data for social passports is collected by managers of educational institutions or by school counselors (a less frequent practice due to the reduced number of social workers at school) and represents a summary of information provided voluntarily by parents. We also asked that social passports should include some additional characteristics of educational institutions that could also help assess their performance: education outcomes (e. g., USE points in Russian language or mathematics), student population, type of institutions, pro-

---

[9] The project and monitoring website: http://www.kpmo.ru/. Before autumn 2014, monitoring data on almost every school in the Russian Federation was publicly accessible. Today, however, the data may only be accessed by registered users or indirectly, via the interface of the Interactive Map of General Education Institutions in Russia (the project website: http://www.educationmap.ru/). The credibility of data uploaded to the system and the validity of indicator calculation methods are widely discussed in the expert community.

[10] There are indicators characterizing financial and economic activities, infrastructure and staff composition of educational institutions, as well as giving some additional information (student population and maximum number of students per class, teacher-student ratio, etc.).

Table 1. **The content of an extended social passport of a school**

Characteristics of the student body	Characteristics of schools
• Proportion of students from multi-child families • Proportion of students from single-parent families • Proportion of students under legal guardianship • Proportion of students from families with both parents unemployed • Proportion of students from single-parent families with the parent unemployed • Proportion of students from families with one or both parents disabled • Proportion of students from families in which both parents have higher education • Proportion of students from families in which at least one parent has higher education • Proportion of students from families living in apartments with few conveniences • Proportion of students from families living in private houses • Proportion of students for whom Russian is not an intrafamilial language • Proportion of students who are adopted • Proportion of students with poor disciplinary records • Proportion of students with criminal records	• Type of educational institution (ordinary secondary education schools, lyceums, gymnasia, specialized schools) • Student population • Number of students per teacher (teacher-student ratio) • Whether the school principal has a professional education (in management) or not • Condition of the building (in critical condition / requires overhaul) • Proportion of first-category teachers • Proportion of higher-category teachers • Proportion of teachers with higher teacher education • Proportion of teachers of retirement age • Urban/Rural school • Average USE points in mathematics • Average USE points in Russian

portion of teachers with specific skills, etc. The key variables from the extended version of a social passport are presented in Table 1.

The approach to using contextual data proposed below is based on essential simplification, as the basic units it analyzes are schools and their aggregate characteristics as opposed to individual students. As a result, some of the effects manifested at a more elementary level of analysis (like the role of family in student achievements, which is an indispensable part of adequate school contribution assessment) are inevitably blurred; that is, they become less statistically distinct when moving to a more aggregate level. However, despite the inherently lower precision as compared to more sophisticated models, the approach can be easily adjusted for school performance analysis, which has been demonstrated in situations when an alternative approach is unavailable due to similar data collection constraints (e. g. [OECD, 2008. P. 15]).

3.1. Analysis through Regression Model Residuals

Having at hand the abovementioned information on a certain (for instance, regional) sample of schools, we can determine the empirical correlation between education outcomes (or any other indicators of school performance function) and contextual characteristics. This correlation is most frequently determined by means of multiple regression analysis. Such analysis involves searching for the optimal

trend line for points scattered over the coordinate plane of a graph (e. g., that of the correspondence between the proportion of highest-category teachers and average USE scores). Multiple regression analysis makes it possible to do the same for an unlimited number of arguments, i. e., for three or more dimensions, such as when it is necessary to describe the variance of the results in relation not only to the proportion of higher-category teachers but also to such characteristics as teacher-student ratio or the proportion of children from low-income families. The analysis is designed to identify empirically the function parameters that provide for an ultimately precise description of the scatter of a specific variable through a number of arguments.

In a simplified vectorial form, the model of this function for the case at hand may be written as follows:

$$Y_i = \beta_0 + B_n \times (CHARACTERISTICS\ OF\ THE\ STUDENT\ BODY)_i +$$
$$+ B_m \times (CHARACTERISTICS\ OF\ SCHOOLS)_i + \varepsilon_i,$$

where Y_i stands for a dependent variable for the ith school. *Any school performance indicator* may be used, but we will use the average USE scores[11] hereinafter to demonstrate the method. Parameter β_0 is a fixed, i. e., independent from the function arguments (characteristics of schools and of the student body) level of USE scores to serve as the basis for all the effects analyzed further; B_n is a vector with n parameters reflecting the degree of correlation between the relevant characteristics of the student body and the average USE points (where n refers to the number of student body characteristics included into the model); B_m is a similar vector with m parameters applied to the characteristics of schools (m stands for the number of characteristics); ε_i is the residual, or "white noise," describing the scatter of USE rates across the schools, which cannot be explained using the rest of the factors included in the model and is treated here as a random effect.

The model presented above implies that all explanatory factors included in it (vectors of characteristics) are related to the independent variable Y_i (education outcomes) linearly, which is quite a realistic

---

[11] A similar procedure may be adapted to identify high-performing and low-performing schools using other academic achievement criteria, as comprehensive assessment of school performance obviously requires an extended list of criteria. The pool of possible criteria can also be formed using the following indicators that may be calculated for individual student cohorts: 1) variability of results on the USE and SFE national tests; 2) the "promoting power" of the school [Balfanz, Legters, 2004] calculated as a proportion of successful graduates in the cohort of students enrolled in 10th grade; 3) the proportion of prize winners in academic competitions and other contests at various levels (or any other indicators of academic or extracurricular performance); 4) the proportion of students continuing education at various levels after graduation, etc.

assumption easily verified through graphic analysis by constructing scatter plots for two variables, one dependent and the other explanatory. Moreover, multiple linear regression has the advantage of allowing one to model nonlinear bonds by using diverse transformations of variables included (e. g., taking logarithms, exponentiation, raising to a power).

All the parameters specified above (β_0, elements B_n and B_m) are evaluated by the ordinary least squares method, which makes it possible to find the parameter values that allow for the most precise description of differences between schools by indicator Y_i using the explanatory variables (characteristics of schools and of the student body). The value of each parameter is interpreted as a simple regression coefficient that shows the average variance of the response variable (education outcomes) if the relevant independent variable increases or decreases by one unit, provided that the rest of the variables in the model remain unchanged (i. e., statistically controlled). Thus, the sign and value of the coefficient demonstrate the nature of correlation (negative/positive, strong/weak). However, regression models do not identify any cause-effect relationships, which means that they do not allow us to state that explanatory values affect the dependent ones; they only make it possible to detect the relationship and determine its nature.

Table 2 shows the values of coefficients (parameters) obtained using multiple regression analysis based on the model described above, using observed data on schools in three regions of Russia in 2011–2012. These results *do not reflect the national situation*, for the sample is limited to three regions and a short time period[12]. *Therefore, we warn against any interventions into the current education policy with reference to the values obtained in this study and stress again that these values are only intended to demonstrate the method in action.*

Table 2 cites only the indicators that showed a stable and statistically significant correlation with the response variable (average USE points in the two major subjects, Russian and mathematics). Therefore, these indicators may be taken as the basis for contextualization—that is, for more reasonable comparisons between educational institutions that consider their current provision of resources and the complexity of their student bodies[13]. In fact, the process of cal-

---

[12] Of course, more advanced models should be ideally tested on the national sample or at least on its representation based on a random or any other representative sample.

[13] As soon as the model demonstrated here describes differences virtually within one cohort of students (data on the USE points is collected for one year), it is certainly an essential simplification. In this situation, the socio-demographic composition of students should be analyzed not with the whole student population but with the relevant cohort, but data contained in social passports of schools makes it impossible (by the way, this problem is eas-

culation consists of a number of iterations that test models with different sets of explanatory variables. These iterations serve to identify the best possible model for each of the dependent variables (USE scores) so that the latter would make it feasible to predict as precisely as possible the differences between schools by the given parameter using a limited list[14] of explanatory variables (these variables should remain maximally orthogonal to each other[15]). Table 2 doesn't provide but parameters of the finite models.

The interpretation of the regression coefficients depends on the scale used to measure explanatory and response variables—whether they were measured in points, proportions, percents, units, dozens of units, etc. As an example of interpretation, we will analyze the "effect" that the "proportion of students from families in which both parents have higher education" has on the average USE scores in mathematics. The coefficient for this variable is 0.06, and it is statistically significant at the level of 99 percent. This means that whenever the proportion of students from families in which both parents have higher education increases by 1 percent, the mean USE score in mathematics increases by 0.06 points on average (by 0.6 points per 10 percent, respectively). When we interpret the coefficient like this, we assume that all the other indicators remain unchanged, as if we were analyzing two absolutely identical schools that only differed in the proportion of students whose both parents had higher education. Likewise, as we can see in Table 2, schools with a 10 percent higher proportion of children from single-parent families perform on average 1 point worse on the USE in mathematics (the negative coefficient of 0.10 multiplied by 10)—again, providing that all the other indicators remain unchanged. Finally, the precision, or explanatory power, of linear regression models is measured through the determination coefficient, which corresponds to the proportion of dispersion of the "dependent" indicator (which is explained by the model) and which is one of the key criteria for choosing the optimal regression model (USE points in two subjects, in our case).

---

ily solved if all characteristics are considered individually). Furthermore, it is also practical in this case to take into account the gender composition of the cohort, which can also have an impact due to the higher average performance of girls [Voyer, Voyer, 2014].

[14] A smaller number of explanatory variables in the model enables analysis of more educational institutions, as not many schools fill out their social passports completely.

[15] One of the problems with modeling bonds by means of multiple regression analysis is the potential statistical relationship (correlation) between the explanatory variables. If the model includes such variables and the relationship between them is rather strong, values calculated for the respective coefficients may be offset, which makes it impossible to determine the exact effects the variables have on fluctuations of the response parameter.

Table 2. **Parameters of regression models describing how schools' average USE scores correlate with characteristics of schools and social composition of the student body**

Parameters of the model	Coefficients for the USE points	
	in mathematics	in Russian
Fixed (not related to explanatory variables) level of the average USE scores (a constant)	42.74** (1.25)	59.61** (1.16)
CHARACTERISTICS OF THE STUDENT BODY		
Proportion of students from single-parent families	−0.10** (0.02)	−0.07** (0.02)
Proportion of students from families in which both parents have higher education	0.06** (0.01)	0.06** (0.01)
Proportion of students for whom Russian is not an intrafamilial language	—	−0.09* (0.04)
Proportion of students with poor disciplinary records	−0.25** (0.09)	−0.25** (0.09)
CHARACTERISTICS OF SCHOOLS		
Municipal school	—	2.06** (0.54)
Lyceum, gymnasium	2.63** (0.62)	3.07** (0.57)
Specialized school	—	1.40* (0.67)
Student population	0.01** (0.00)	0.01** (0.0)
Student-teacher ratio	−0.24** (0.06)	−0.27** (0.06)
Proportion of higher-category teachers	0.06** (0.02)	0.05** (0.01)
Proportion of first-category teachers	0.06** (0.02)	0.05** (0.01)
PRECISION (EXPLANATORY POWER) OF MODELS		
Determination coefficient (R^2)	0.27	0.35

Note: Round brackets are used to specify standard error of regression coefficients; *—significant at the level of 5%; **—at the level of 1%.

Overall, the analysis has shown that academic performance assessed by the USE points in Russian and mathematics differs consistently among schools with different social composition of students: a more favorable social environment yields higher academic performance. The most sensitive environment characteristics in our sample include: level of parental education, proportion of students from single-parent families, and presence of especially difficult students (children with disciplinary or criminal records, etc.). Regardless of their competencies, teachers find it more "comfortable" to work with students whose parents express concern about the achievement of their

children and invest actively in various extracurricular activities than with students whose parents cannot afford to do these things. Deviating behavior, cultural barriers, and possible lack of parental attention are things that cause troubles for teachers that must be considered when assessing their performance and the performance of the schools they work in.

Alongside the characteristics of the social composition of students, a number of other characteristics correlate with the average performance rate, providing for more qualitative differentiations between schools. The most meaningful of these characteristics include belonging to one of the special categories of educational institutions (lyceum, gymnasium, or specialized school), size of the school (student population), and parameters of staffing resources (proportion of higher-category teachers). Naturally, these factors often overlap: lyceums and gymnasia are normally large institutions with high proportions of expert teachers. This is, in fact, the most privileged category of schools, and requirements for their education outcomes should reasonably be as high as possible. However, each of the three factors reveals an independent correlation with the performance rate. Thus, when comparing performance of schools of the same type with similar staffing resources, one cannot ignore the distinction in their size, which may conceal, for example, differences in the amount of financing[16].

The models differ somewhat for the two dependent variables (the USE points in mathematics and Russian). First, the coefficient values are different (but not the sign, or nature, of the relationship). Second, the model for the USE points in Russian looks slightly more detailed, as it includes factors such as the status of school in the locality and the proportion of students for whom Russian is not an intrafamilial language. Without going into a detailed explication of possible reasons for such differences (we partly dwelled on them in [Yastrebov et al., 2013]), we will simply point out that regression analysis is a rather flexible tool that facilitates determining differences in the sensitivity of performance rate indicators to various parameters of conditions, resources and social context. This property makes it possible to contextualize the very tool of contextualization, depending on the conditions, samples, and sets of variables to which it is applied.

Any assessment based on direct comparison of USE scores that does not consider the social context and conditions of school operation is invalid, which is well illustrated by the graphs in Figure 1. The horizontal axis displays observed USE points in the relevant subject, while the vertical axis displays USE points approximated through the parameters of the model presented in Table 2 (these are contextualized estimates, i.e., estimates of expected USE scores calculat-

---

[16] Direct information about the financing of educational institutions was not used here, but it can also be considered in construction of similar models.

Figure 1. **Variation of average school USE scores and their contextualized points, 2011**
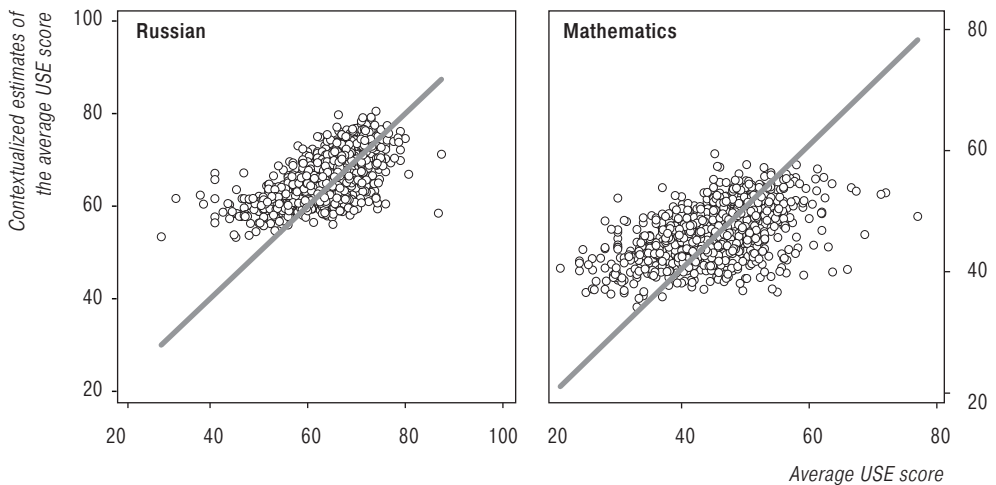

ed with the social context and school characteristics considered). As for the contextualized estimates, they are calculated for each school using ordinary formulae in which expected USE score values are obtained by summing up the related effects (refer to analysis results in Table 2):

Expected/contextualized index of the average USE score in mathematics = 36.5 + 2.3 (for lyceums and gymnasiums) + + 0.007 × *"student population"* + 6.1 × *"proportion of higher-category teachers"* − 8.4 × *"proportion of students from single-parent families"* + 5.5 × *"proportion of children from families in which both parents have higher education"* − 0.25 × *"proportion of children with criminal or disciplinary records"*

Expected/contextualized index of the average USE score in Russian = 59.61 + 2.06 (for municipal schools) + 3.07 (for lyceums and gymnasiums) + 1.4 (for specialized schools) + 0.01 × *"student population"* − − 0.27 × *"student-teacher ratio"* + 0.05 × *"proportion of higher-category teachers"* + 0.05 × *"proportion of first-category teachers"* − − 0.07 × *"proportion of children from single-parent families"* + + 0.06 × *"proportion of children from families in which both parents have higher education"* − 0.09 × *"proportion of children for whom Russian is not an intrafamilial language"* − 0.25 × *"proportion of children with criminal or disciplinary records"*

Of course, the expected USE score values calculated with these formulae should be interpreted with special care, considering the statistical error implied by coefficient calculation (see standard error values

Figure 2. **Distribution of regression model residuals for the USE scores in Russian and mathematics**

Residual values for the model using the criterion of USE points
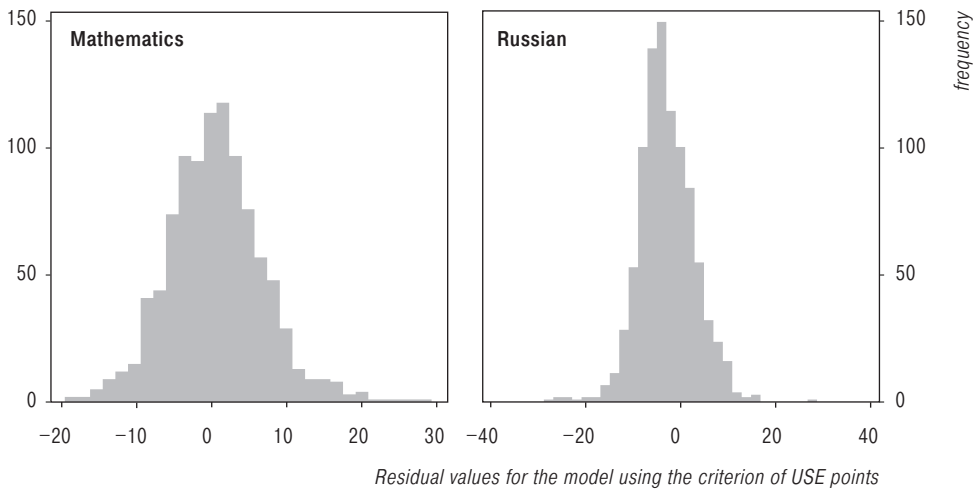
in Table 2). Their routine application only serves to provide a reference point that can be compared with the observed results. In this case, the resulting difference (the so-called residual) shows by how much the observed performance rate indicators are higher or lower than the level expected of the specific school given its resources and context.

As we see in Figure 1, observed and contextualized points differ overwhelmingly: their perfect match is displayed by a slope where the abscissa equals the ordinate at each point. Points lying above the slope correspond to schools whose contextualized outcomes turned out to be higher than the observed values (potentially, these are the schools that need to be developed), while points lying below account for schools who had higher USE scores than the acceptable values determined with due regard to the complicated social context (these are the schools that successfully overcome the inherent restraints and are potentially efficient).

Alternatively, the nature of these deviations, as well as the level of relative performance of educational institutions may be identified through analysis of regression model residuals, which are the difference between the observed average USE score for a specific school and the value obtained by modeling with due regard to contextual factors.

Figure 2 demonstrates the observed distribution of these regression residuals. The distribution is close to the Gaussian curve, with the mathematical expectation almost equaling zero (one of the key requirements to quality specification of regression models estimated with the least square method). The values of residuals show by how

much outcomes of specific schools deviate—in a positive or negative way (depends on the sign)—from the ones predicted by the model using the factors considered. Hence, the residuals can be interpreted as *absolute over- or underperformance by productivity rate indicators* (expressed in corresponding units; in this case, in average USE points) *as compared to the estimated (statistically average) results for schools with identical sets of resource constraints and student body characteristics.* This approach obviously sets higher expected outcome standards for schools operating in more favorable contexts.

Technically, the regression model treats residuals as random phenomena, but this can only be true if the model embraces the entire range of factors describing the distribution of the parameter in question. This is obviously not true in our case: the analysis ignores data on the content of curricula, on students' personal skills, etc. This is why residual analysis may well be regarded as a self-sufficient subject of study. Having analyzed these deviations, we can conveniently divide all schools into two groups: *resilient schools* (capable of overcoming inherent constraints and demonstrating higher results than those prescribed by the pre-set model values) and *failing schools* (demonstrating much lower results than those prescribed by the pre-set model values). The distribution patterns in Figure 2 display these groups through positive (to the right of zero) and negative (to the left) residual values.

Thus, the degree to which the observed performance indicator values deviate from the ones estimated based on the context and conditions can be regarded as a marker of the potential efficiency or inefficiency of an educational institution. Critical values for these deviations may be set randomly; however, given the inevitable errors in calculation of expected values by means of regression analysis, it is also desirable to calculate the confidence intervals of such values for each educational institution. In particular, the range of permissible values can be determined by constructing formulae for the upper and lower thresholds that would not be based on regression coefficients but rather would use the upper and lower limits of their confidence intervals. If the observed values fall within the determined thresholds, it will mean that the school generally demonstrates results typical (i. e., statistically undistinguishable from the mean values) for schools with similar characteristics. Conversely, if observed USE scores are lower (or higher) than the specified threshold, we can say with high confidence that the school shows substantially lower (or higher) results than expected on the assumption of its contextual and other characteristics.

3.2. Analysis by means of the index of school social well-being

There is another way of using regression analysis data to consider social context when assessing education outcomes at schools—by means of the conventional index of school social well-being (ISSWB). It is simpler and more illustrative, but these benefits are associated with some drawbacks that will be described below.

Table 3. **Values of regression equation coefficients as affected by characteristics of the social composition of students and recommended specific weights in the overall index of school social well-being**

	For equations with dependent variable "average USE points in mathematics"	For equations with dependent variable "average USE points in Russian"	Recommended specific weight in the overall index structure
Proportion of students from single-parent families	−0.10	−0.07	~20%
Proportion of students from families in which both parents have higher education	+0.06	+0.06	~15%
Proportion of students with disciplinary records	−0.25	−0.25	~65%
The sum of absolute values of coefficients	0.41	0.38	(100%)

Student body characteristics consistently correlated with USE scores include the following parameters (see Table 2):

1) proportion of students from single-parent families (−);
2) proportion of students from families in which both parents have higher education (+);
3) proportion of students with criminal or disciplinary records (−).

We do not mention the "proportion of students for whom Russian is not an intrafamilial language" here because it was only relevant for one of the models, so we will content ourselves with the three parameters specified above in order to avoid multiplying extra data entities.

As long as these characteristics are measured in the same units (percents) in our regression model, we can easily use the coefficients to determine the specific weight of each characteristic within the aggregate index of school social well-being. Table 3 gives again the values of these coefficients obtained by regression analysis. It is worth recalling that the coefficients can be regarded as characterizing the contribution made by specific parameters in education outcomes of the school. This information provides the basis for assessment of the approximate specific weight of contextual indicators in the ISSWB, which would allow for an integrated description of the social composition of students that can be used to distinguish among educational institutions to ensure a more adequate comparison.

Table 4. **Correlation of social context indicators with those of performance**

Social context indicators	Average USE points in Russian	Average USE points in mathematics
ISSWB	0.50	0.44
Proportion of students from single-parent families	−0.21	-0.22
Proportion of students from families in which both parents have higher education	0.44	0.38
Proportion of students with criminal or disciplinary records	−0.38	−0.33
Proportion of students for whom Russian is not a native language	−0.11	−0.09

Note: all coefficients are significant at the level of 1%

Table 5. **Grouping of urban schools by the ISSWB values**

ISSWB level (by quintiles)	Interval of observed ISSWB values	Average USE points in Russian	Average USE points in mathematics
Low (quintile 1)	[0; 79.5]	58.1	40.7
Below average (quintile 2)	(79.5; 81.6]	59.1	42.9
Average (quintile 3)	(81.6; 83.9]	62.6	44.4
Above average (quintile 4)	(83.9; 86.4]	63.0	45.6
High (quintile 5)	(86.4; 100]	64.2	46.7

The approximate specific weights for the index components can be easily calculated by dividing the absolute values of individual coefficients by their total sum, for example. Table 3 determines recommended values obtained by averaging the parameters of the two models (for the USE scores in Russian and in mathematics). Therefore, one conceivable formulae of the index may be as follows:

ISSWB = 85 + 15 × *"proportion of students from families in which both parents have higher education"* − 20 × *"proportion of students from single-parent families"* − 65 × *"proportion of students with criminal or disciplinary records,"*

where the first value (85) only serves to bring the index to the scale 0–100. It becomes perfectly clear that the index equals 100 in the most favorable situation for the school (when all positive characteristics of the social composition are 1 and all negative ones are zero)

and 0 when the school is the most vulnerable. Other methods of "convoluting" the integrated index may also be used (for example, the Z-standardization widely applied in statistics). The paramount requirement is that the scale should consider various weights of constituent parameters that correspond to their potential contribution in the variance of performance rate indicators. The ISSWB is calculated for each educational institution based on the source data contained in the social passport.

Performance indicators are more sensitive to the ISSWB than to its individual components, which is proved by simple correlation analysis of data in the same three regions for which we calculated regression model parameters (Table 4).

We will now demonstrate how the ISSWB may be used to compare the performance of educational institutions with due regard to their social context.

The principle of its application is largely the same as that of the method of considering contextual information used in NAPLAN: it is necessary to distinguish groups of "statistical neighbors," i. e., educational institutions with similar conditions, to perform comparisons within these groups. Yet, other qualitative differences between schools (besides those measured with the ISSWB) should be determined in advance. Let us suppose that such differences include the type of institution (ordinary secondary general education school, lyceum, gymnasium, or specialized school) and its urban or rural status. We suggest taking a group of ordinary urban secondary general education schools for illustrative purposes.

Within this group, schools may also differ greatly by the social composition of their students, so next we should differentiate among them using the index constructed above. To do this, we can divide all schools into five quintiles, i. e., equally numerous subsets with the ISSWB values aggregated bottom-up, where the lowest quintile includes schools with the least advantaged student populations and the highest quintile includes those with the most advantaged student populations. Table 5 provides some parameters for these groups based on the aggregate total of urban schools in three regions; namely, the interval of observed ISSWB values and the average USE scores in two fundamental subjects (the list of characteristics may also be extended to include other indicators of educational institution efficiency).

As we see in Table 5, the average level of scores differs consistently and (more importantly) predictably across the groups of urban schools differentiated according to their levels of social well-being. Nevertheless, as you will see below, even these relatively homogeneous groups demonstrate differences in outcomes. However, it is only now, after considering the essential conditions of performance of educational institutions, that we should regard these differences as indicators of potential efficiency or inefficiency.

Figure 3. **Distribution of USE scores in two fundamental subjects within the group of urban schools with high ISSWB values**
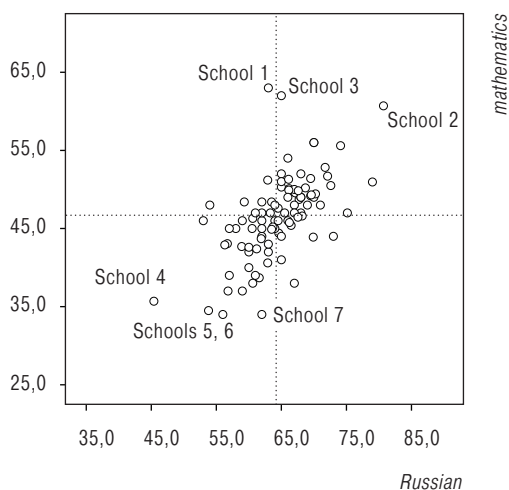
Note: The horizontal and vertical lines in the graph mark the values of the group's average USE scores in mathematics (46.7) and in Russian (64.2), respectively

Figure 3 shows the distribution of urban schools from quintile 5, the most favorable in its social composition, in two dimensions: the average USE scores in Russian and in mathematics. Each point on the graph corresponds to one school. As we can see, the largest part of educational institutions is concentrated in the area of average education outcomes for the group which can be considered typical for schools with relevant qualitative characteristics (the area is circled by an ellipse). The points far beyond the ellipse boundaries tell us that, for some reason, the respective schools demonstrate results too high or, conversely, too low for their group. These are the schools that should become the focus of special attention of education authorities and, perhaps, the targets of intervention.

Thus, for example, the school located in the upper right quadrant, denoted as School 2, deserves special attention: its performance both in Russian and in mathematics is considerably higher than the average group values, while formal parameters of context and conditions are essentially the same for all schools. This phenomenon may have a number of explanations, from a properly organized learning process, specific features of the school's curriculum and high-quality teaching to deliberate (or unintentional) bias[17]. In any case, such

---

[17] The idea of bias, unfortunately, cannot be excluded in light of the notorious outcomes of the 2013 USE campaign.

performance rate indicators should encourage education authorities to investigate deeper into the situation at the school and to make relevant managerial decisions (for example, by studying the school experience in detail with the intent of applying the conclusions to in other educational institutions).

The lower left quadrant in Figure 3 contains schools (School 4 and School 5) whose low performance cannot be justified by specific conditions or context of their operation because the sample group has already been "leveled out" with regard to these parameters. This is why targeted education policy on these schools should focus primarily on identifying the reasons for which they are incapable of providing the acceptable level of education quality.

The proposed analysis method also enables the identification of various borderline cases, like with Schools 1 and 3, where relative over- or underperformance only takes place for a specific group of indicators. The values of the USE scores in two fundamental subjects and the illustration in Figure 3 are simply used as examples to demonstrate the general principle and prospects of the method of considering contextual information to compare the performance of educational institutions. The range of the indicators can and should be extended to include the comprehensive list of markers to enable the monitoring of how well schools balance their education functions and socialization. This analysis, just as regression residual analysis, would be more accurate if completed with dynamic comparisons, which take into account how indicators change over a relatively long period of time, because otherwise it is impossible to evaluate the consistency (as opposed to random nature) of the differences. While performing these integrated comparisons, it is important to consider the constraints imposed on implementation of relevant functions by the social context and resources available to the school.

The overall principle of analysis described here reproduces the logic of regression residual analysis: first we calculate typical performance rate indicators for schools with identical characteristics, and then we compare them to the real indicators of specific educational institutions. To be fair, division of schools into groups of "statistical neighbors" is little different from the cluster method which has already been used by some researchers in Russia [Agranovich et al., 2008; Konstantinovsky, Vakhshtayn, Kurakin, 2013] and which has its own constraints. In particular, this approach ignores the issue with borderline ISSWB values (see, for instance, interval borders in Table 5) and other continuous indicators that serve as the basis for grouping, as observed deviations in outcomes (including those in Figure 3) can be explained, inter alia, by certain differences within the groups. Therefore, we believe the first method is more preferable.

Conclusion   Russia has not yet developed the tradition of considering contextual data in its systems of education quality management and school performance assessment. A number of studies conducted in Russia and abroad, including this article, demonstrate why disregard for such data is absolutely indefensible, especially in terms of developing an efficient and balanced education policy that could find custom solutions to emerging problems.

Tools and methods of using contextual data in education management are quite diverse and are not limited to the approaches described above. Unfortunately, there are no critically important prerequisites for the development of such tools and their broad application. These prerequisites include consistent collection of data on social composition of students, which is rather realizable today, given the experience of introducing social passports for educational institutions in a number of Russian regions. However, even when this information is accessible, it remains inaccurate and disconnected, as data is provided on a voluntary basis and is not subject to any accepted standards of accounting and primary processing. Still, there is a more global problem impeding the use of contextual information in education management, which will prove difficult to solve in the foreseeable future: the universal lack of trust in statistical data submitted by school community members and education authorities, which inevitably impairs the efficiency of using such data for management purposes. The USE statistics, again, are a prominent example. Many experts believe these statistics are poorly associated with real curricula and are extremely vulnerable to fraud [Bolotov, Valdman, 2013].

The problems at hand should be solved, as this is quite literally the clue to enhancing the competitive power of the Russian education system in the global race for high education standards, where Russia's position is still unstable, as international research proves. While leading or successfully competing with many developed countries in some studies (PIRLS, TIMSS), Russia is remarkably weaker in others (PISA)[18]. The best short-term policy would probably be to focus on the problems that are easy and inexpensive to solve. One such challenge is the introduction of a unified procedure of collecting detailed contextual information in all schools of the Russian Federation. This could be realized as part of the education system monitoring carried out under the education initiative "Our New School" (with modifications to the scope of data collected, of course) or on the basis of such ambitious but helpful initiatives as the "Open Government" project. Experts from regional education quality assessment centers and researchers in the field should be involved in the exchange of ideas on the methods to be used, including discussion of possible models

---

[18] Basic PISA 2012 results. http://www.centeroko.ru/public.htm

to assess performance of educational institutions with due regard to contextual data (including the efforts of the Center for Social and Economic Development of Schools under the NRU‑HSE Institute of Education). This array of information would allow education authorities to have a comprehensive idea of the situation in Russia's system of secondary education.

The future, however, will inevitably require a transition to a more precise accounting system that will record performance dynamics, academic achievements and important socio-demographic information on each student, including their possible transfers within the education system. Such system will provide access to the most advanced models of education institution assessment performance applied by developed countries today. As this step involves considerable financial investment and requires removal of some legal restrictions, some preparatory work should already be started today, including expert discussions of possible monitoring methods and pilot projects in some of the regions. We believe that this could be done within the framework of the Education Development Federal Targeted Program for 2011–2015 and the State Program "Education Development" for 2013–2020.

In summary, the All-Russian system of education quality assessment faces the following urgent objectives:

- include contextual data in the scope of information collected;
- develop models and tools for contextual data aggregation, including calculation of "national indexes";
- introduce methods of taking contextual data into consideration, including monitoring, in analysis of assessment results;
- integrate standards of contextual data representation in communications systems (websites, public reports).

It is only when simple but crucial steps are taken to include the context of education institutions in assessment algorithms that we will be able to talk about education quality management—not faulty management responding to external signals as a reflex, but focused management, based on analysis of high-quality data and the implementation of targeted strategies declared by national education policy[19], which the international community refers to as "raising the bar and closing the gap."

---

[19] You will find more details on practices of such management, as well as on the relevant programs and projects implemented at national and regional levels and widely used abroad, in the collection of materials on the subject [Frumin et al., 2012].

References

1. Agranovich M. (2008) Indikatory v upravlenii obrazovaniem: chto pokazy-vayut i kuda vedut? [Indicators in Education Management: What Is Indicated and What's the Purpose?]. *Voprosy obrazovaniya*, no 1, pp. 120–145.

2. Agranovich M. (2010) Otsenka effektivnosti deyatelnosti regionalnykh organov ispolnitelnoy vlasti v sfere obrazovaniya [Assessing Performance of Regional Executive Authorities in Education]. *Zhurnal rukovoditelya upravleniya obrazovaniem*, no 6, pp. 39–50.

3. Agranovich M., Alekseeva A., Zaytseva O., Kogan Y., Matveeva N., Mitrofanov K., Prudnikova V., Selivyorstova I. (2008) *Upravlenie kachestvom obrazovaniya v regione na osnove pokazateley i indikatorov: sb. instruktivno-metodicheskikh materialov* [Managing Education Quality in Regions Based on Factors and Indicators: Guidelines and Procedures], Moscow: Logos.

4. Aitkin M., Longford N. (1986) Statistical Modeling Issues in School Effectiveness Studies. *Journal of Royal Statistical Society, series A*, vol. 149, no 1, pp. 1–43.

5. Balfanz R., Legters N. (2004) *Locating the Dropout Crisis—Which High Schools Produce the Nation's Dropouts? Where Are They Located? Who Attends Them?* John Hopkins University CSOC Report. Baltimore, MD.

6. Barr R., Dreeben R. (1983) *How Schools Work*. Chicago: University of Chicago.

7. Bochenkov S., Valdman I. (2013) Interpretatsiya i predstavleniye rezultatov YGE: problemy i vozmozhnye resheniya [Interpretation and Representation of USE Results: Problems and Possible Solutions]. *Voprosy obrazovaniya*, no 3, pp. 5–24.

8. Bolotov V., Valdman I. (2013) Vidy i naznacheniye programm otsenki rezultatov obucheniya shkolnikov. *Pedagogika*, no 8, pp. 15–26.

9. Burstein L. (1980) Issues in the Aggregation of Data. Review of Research in Education, vol. 8, pp. 158–236.

10. Dyer H. S., Linn R. L., Patton M. J. (1969) A Comparison of Four Methods of Obtaining Discrepancy Measures on Observed and Predicted School System Means on Achievement Tests. *American Educational Research Journal*, vol. 6, no 4, pp. 591–605.

11. Frumin I., Kosaretsky S., Pinskaya M., Grunicheva I., Timkova T. (eds) (2012) *Vyravnivaniye shansov detey na kachestvennoye obrazovaniye: sb. materialov Komissii Obshchestvennoy palaty Rossiyskoy Federatsii po razvitiyu obrazovaniya* [Balancing Children's Chances for Quality Education: Collection of Materials Issued by the Education Development Committee of the Civic Chamber of the Russian Federation], Moscow: National Research University—Higher School of Economics.

12. Hanushek E. A. (1986) The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature*, vol. 24, no 3, pp. 1141–1177.

13. Hanushek E. A. (1989) The Impact of Differential Expenditures on School Performance. Educational Researcher, vol. 18, no 4, pp. 45–62.

14. Konstantinovsky D., Vakhshtayn V., Kurakin D. (2013) *Realnost obrazovaniya. Sotsiologicheskoe issledovanie: ot metafory k interpretatsii* [Education Reality. Sociology Research: From Metaphor to Interpretation], Moscow: Center for Social Forecasting and Marketing.

15. Konstantinovsky D., Voznesenskaya Y., Cherednichenko G., Khokhlushkina F. (2011) *Obrazovaniye i zhiznennye trayektorii molodyozhi: 1998–2008 gody* [Education and Life Trajectories of Youth: 1998–2008], Moscow: Institute of Sociology, Russian Academy of Sciences.

16. Levin H. M. (1994) Production Functions in Education. *International Encyclopedia of Education* (eds T. Husen, T. N. Postlethwaite), New York: Pergamon, pp. 4059–4069.

17. McCall M.S., Kingsbury G. G., Olson A. (2004) *Individual Growth and School Success*. Lake Oswego, OR: Northwest Evaluation Association.

18. OECD (2008) *Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-Added of Schools*. OECD: Paris.

19. Pinskaya M., Kosaretsky S., Frumin I. (2011) Shkoly, effektivno rabotayushchie v slozhnykh sotsialnykh kontekstakh [Schools Working Effectively in Complicated Social Contexts]. *Voprosy obrazovaniya*, no 4, pp. 148–177.

20. Prakhov I., Yudkevich M. (2012) Vliyaniye dokhoda domokhozyaystv na rezultaty YGE i vybor vuza [How Household Income Affects the USE Scores and the Choice of Higher Educational Institution]. *Voprosy obrazovaniya*, no 1, pp. 126–147.

21. Roshchina Y. (2012) Semeyny kapital kak faktor obrazovatelnykh vozmozhnostey rossiyskikh shkolnikov [Family Capital as a Factor of Educational Opportunities for Russian School Students]. *Voprosy obrazovaniya*, no 1, pp. 257–271.

22. Rumberger R. W., Palardy G. J. (2004) Multilevel Models for School Effectiveness Research. *Handbook of Quantitative Methodology for the Social Sciences* (ed. D. Kaplan), Thousand Oaks, CA: Sage, pp. 235–258.

23. Rumberger R. W., Thomas S. L. (2000) The Distribution of Dropout and Turnover Rates Among Urban and Suburban High Schools. *Sociology of Education*, vol. 73, no 1, pp. 39–67.

24. Shavelson R., McDonnell L., Oakes J., Carey N. (1987) *Indicator Systems for Monitoring Mathematics and Science Education*. Santa Monica, CA: RAND.

25. Sobkin V., Pisarsky P. (1998) *Tipy regionalnykh obrazovatelnykh situatsiy v Rossiyskoy Federatsii* [Types of Regional Educational Situations in the Russian Federation], Moscow: Center for Sociology of Education, Russian Academy of Education.

26. Uvarov A., Yastrebov G. (2014) Sotsialno-ekonomicheskoe polozhenie semey i shkola kak konkuriruyushchie faktory obrazovatelnykh vozmozhnostey: situatsiya v Rossii [Socioeconomic Status of Families and School as Competing Factors of Educational Opportunities: The Situation in Russia]. *Mir Rossii: Sotsiologiya, etnologiya*, no 2, pp. 103–132.

27. Valdman I. (2013a) Osobennosti organizatsii monitoringov uchebnykh dostizheniy v stranakh mira: natsionalnaya programma otsenki NAPLAN v Avstralii [Specifics of Student Performance Monitoring in Countries of the World: The National Assessment Program—Literacy and Numeracy (NAPLAN) in Australia. *Problemy sovremennogo obrazovaniya*, no 2, pp. 5–18.

28. Valdman I. (2013b) Osobennosti organizatsii monitoringov uchebnykh dostizheniy v stranakh mira: natsionalnaya programma otsenki SIMCE v Chili [Specifics of Student Performance Monitoring in Countries of the World: The Education Quality Measurement System (Sistema de Medición de la Calidad de la Educación SIMCE) in Chile]. *Problemy sovremennogo obrazovaniya*, no 5, pp. 133–149.

29. Voyer D., Voyer S. D. (2014) Gender Differences in Scholastic Achievement: A Meta-Analysis. *Psychological Bulletin*, vol. 140, no 4, pp. 1174–204.

30. Willms J. D. (1992) *Monitoring School Performance: A Guide for Educators*. Washington, DC: Falmer.

31. Willms J. D., Raudenbush S. W. (1989) A Longitudinal Hierarchical Linear Model for Estimating School Effects and Their Stability. *Journal of Educational Measurement*, vol. 26, no 3, pp. 209–232.

32. Yastrebov G. (2010) Investitsii v chelovecheskiy kapital (Effekt kulturnoy preemstvennosti vs effect dokhoda) [Investments in Human Capital (Cultural Continuity Effect vs. Income Effect]. *Obshchestvennye nauki i sovremennost*, no 2, pp. 70–77.

33. Yastrebov G. (2012) Chto my znayem o roli shkol v vosproizvodstve sotsialnogo neravenstva (po materialam vedushchikh sotsiologicheskikh izdaniy nachinaya s 1970-kh godov) [What We Know about the Role School Plays in Reproducing Social Inequality (based on the major sociology periodicals beginning from the 1970s)]. *Vyravnivaniye shansov detey na kachestvennoye obrazovaniye: sb. materialov Komissii Obshchestvennoy palaty Rossiyskoy Federatsii po razvitiyu obrazovaniya* [Balancing Children's Chances for Quality Education: Collection of Materials Issued by the Education Development Committee of the Civic Chamber of the Russian Federation], Moscow: National Research University—Higher School of Economics, pp. 83–99.

34. Yastrebov G., Bessudnov A., Pinskaya M., Kosaretsky S. (2013) Problema kontekstualizatsii obrazovatelnykh rezultatov: shkoly, sotsialny sostav uchashchikhsya i uroven deprivatsii territoriy [The Problem of Academic Performance Contextualization: Schools, Social Composition of Students, and the Extent of Deprivation of Territories]. *Voprosy obrazovaniya*, no 4, pp. 188–246.