

BULGARIA

Pilot implementation of statistical models for estimation of the value-added of Bulgarian schools using national student assessment data

Key results and findings

June 2013

**WORLD BANK
EUROPE AND CENTRAL ASIA REGION
HUMAN DEVELOPMENT UNIT**



Acknowledgements

This report was prepared under the programmatic Education Sector Knowledge and Advisory Services Program of the World Bank (WB) in response to the request of the Bulgarian Ministry of Education Youth and Science (MEYS) to pilot and document the implementation of a set of statistical models for estimation of the value-added of Bulgarian schools using data from the national census-based student assessments. The pilot and the report were prepared by a team led by Plamen Danchev (Education Specialist, World Bank) and comprising of the following researchers from the Bulgarian Association for Educational measurement and Evaluation (BAEME): prof. Kiril Bankov (Lead Consultant, BAEME), Vessela Stoimenova (Consultant, statistical analyses, BAEME), Dimitar Atanassov (Consultant, statistical analyses, BAEME)). The report was prepared under the guidance of Alberto Rodriguez, Education Sector Manager for Europe and Central Asia, World Bank.

The scope of the pilot implementation, the key parameters for the activity and the required data for the analysis were jointly defined by the WB and a working group of MEYS experts led by Mr. Krassimir Valchev, Secretary General of the MEYS and comprising Evgenia Kostadinova (Director for Curriculum and School Programs, MEYS), Orlin Kouzov (Director for IT in Education, MEYS), Neda Kristanova (Director, CKOKUO -National Testing Agency), Svetla Petrova (Acting Director, CKOKUO), Sashko Arabadjiev (Expert, CKOKUO), Dimitar Enchev (Expert, MEYS), Lyubka Grueva (Expert, MEYS), Rumiana Tomova (Expert, MEYS), Stela Mitsova (Expert, MEYS) and Radosveta Drakeva (Adminsoft, developer of the Education Management Information System of the MEYS).

Contents

EXECUTIVE SUMMARY	2
INTRODUCTION	6
DATA USED FOR THE PILOT IMPLEMENTATION OF THE SCHOOL VALUE-ADDED MODELS	7
DATA PROCESSING.....	8
DESCRIPTIVE STATISTICS	9
REGRESSION BASED MODELS	11
IMPACT OF CONTEXTUAL VARIABLES REPRESENTED BY REGRESSION LINES	18
CORRELATION BETWEEN THE TEST SCORE AND THE VALUE-ADDED	18
KEY FINDINGS	19
PROPOSED NEXT STEPS AND AREAS FOR FURTHER ANALYSIS.....	20
ANNEX 1. QUESTIONNAIRE: STUDENTS BACKGROUND CHARACTERISTICS.....	21
ANNEX 2. VALUE-ADDED ESTIMATES FOR THE SCHOOLS IN A BULGARIAN MUNICIPALITY	25

Executive Summary

Background

1. The objective of this report is to describe the process and results from the pilot implementation of statistical models for measuring the value-added of Bulgarian schools using national student' assessments data. The report is intended to present the technical aspects of the pilot (in the context of Bulgaria student assessment framework and the available data used) and to stimulate the technical level discussions on the relevance and applicability of the piloted models prior to tabling the value-added approach for public discussions and policy decisions.
2. The pilot implementation documented in the present report is a result of the ongoing dialogue between the Ministry of Education, Youth and Science (MEYS) and the World Bank (WB) on strengthening the student assessment system and improving the use of assessments data. The background for this pilot was set by the World Bank's report "*Bulgaria: Using Student Assessment data for Estimating the Value-added of Schools - Techniques and Application for School Improvement and Accountability*" (April, 2011).

The meaning of "school's value added measure"

3. The school's value added represents a measure of school performance derived from statistical analysis of the raw test scores of one and the same cohort of students in at least two points in time, further adjusted to control for the socio-economic characteristics of the students. Through such analyses the students' achievement could be decomposed into components attributable to schools and components attributable to the students' background characteristics. International experience suggests that the school value-added measures (VAMs), based on a strong and reliable student assessment framework, are arguably the most precise and objective indicators for school performance developed to date. VAMs are gaining in popularity and many high-performing educational systems use them as part of their school improvement or accountability frameworks.

Relevance of school's value added measures in the context of the Bulgarian education system

4. Analyzing the test results through value-added analysis is a new approach for Bulgaria and the findings from this pilot are intended to inform both technical level experts and decision makers at the MEYS about the advantages and limitations of VAMs in Bulgaria. Since 2007, when external standardized tests were first introduced in Bulgaria, the national student assessment has expanded in coverage and presently all national assessments are census based (covering all students in the tested cohorts) and administered annually at the end of elementary (Grade 4), primary (Grade 7) and upper secondary (Grade 12, Matura) levels. The new draft law on Preschool and School Education envisions a new lower secondary education stage covering Grades 7-10 and additional national census-based annual assessment at the end of Grade 10. However, the assessment results have not yet been used effectively for accountability and school improvement purposes and for incentivizing school performance gains. Presently, the national student assessment results in Bulgaria are reported on a non-transformed raw point scale accompanied by a table that allows a transformation of the raw points to the six-grade rating scale, adopted for measuring students' performance. Based on this, each school receives an average score and some means of comparison with other schools in the country. Datasets with school results are not publicly available to avoid improper comparisons and conclusions.
5. While raw test scores from the national assessment are sufficient to identify high and low performing schools, their use for accountability and awarding high performance is perceived as problematic. At the same time the recently completed new draft Law on Preschool and School Education envisions additional performance-based financing whereby best performing schools would receive performance awards while low-performing schools would be entitled to a targeted support for improving students' educational outcomes. Identifying appropriate indicators against which performance of schools will be measured to guide the decisions for allocation of the additional performance-based funding streams becomes central to the implementation of the envisioned policy. The raw test scores mirror the composition of the school cohorts in terms of student background characteristics and using them as an indicator for school contribution and performance is deemed unfair in the context of the Bulgarian highly selective education system, where the combination of free school choice and early selection of students into "good" and "bad" schools produces significant between school variance in terms of student achievements. The need for the development of fair and objective measures for school performance that isolate the contribution of schools to students achievement from the factors over which schools have limited or no control (e.g. the socio-

economic status and family or ethnic background of the students) has stimulated discussions among education policy-makers and stakeholders on the need to develop models for estimating the value-added measures of school performance.

Data used for the pilot implementation

6. The Bulgarian pilot implementation of the value added models benefit from the availability of student level data for the first cohort of students that took both the national assessment test in Grade 4 (back in 2009) and the national assessment tests in Grade 7 (2012). This has enabled the use of the tests scores in these two points in time in order to implement the value-added analysis. The subjects with stable and common subject base across the education stages used for the pilot are Bulgarian language and literature and Mathematics. A total of 1918 schools are included in the analysis based on a full dataset for approximately 48,500 students from the analyzed cohort (compared to an estimated averaged total of around 60,000 students).
7. The screening of readily available and routinely collected data on student and school level characteristics revealed that the only available data that could be used for value-added modeling are the students' age, gender, the language spoken at home, complemented by data on the different types of schools. The analyzed cohort comprises of nearly 80% speaking Bulgarian at home, followed by Turkish, Roma and "others". The female students exceed by 3% the male students. The linkage of the students and their tests scores in the subjects in grades 4 and 7 is made through anonymized unique substitutes for the students' national social security ID number (called EGN). This ID was instrumental for eliciting the information about the date of birth. The Grade 4 national assessment data for the students tested back in 2009 contains, in addition to the test scores, information about the language spoken at home.
8. A joint World Bank-MEYS review of the student level data and student background characteristics concluded that the value-added analysis would benefit from including more individual level information about the students. For this reason, a questionnaire was designed and fielded to collect additional information for all students of the cohort. The questionnaire contains questions on the language spoken at home (to confirm the information collected in 2009), the members of the family, the education of the parents, and information intended to build a profile for the socio-economic status of the students, such as employment status of the parents, the number of the books at home, etc. The full version of the questionnaire is included in Annex 1 to this report. The administration of the questionnaire returned a very low response rate, triggering a decision to not use the data from the questionnaire and to pilot the models using the readily available data described earlier.

Missing Data and Inconsistencies

9. Beside the suboptimal range of student background characteristics, the dataset for the pilot value added modeling suffered from some missing data and data inconsistency. The number of children who have participated in Grade 4 assessment is 59,288, while the number of the children measured by Grade 7 national assessment is 60974. However, after merging the datasets and linking data through the unique student IDs, the total number of students with full set of data fell to 48,529. Based on additional data analysis the hypothesis for a large number of drop outs was ruled out and the discrepancy is explained by the fact that apparently not all students had taken exams on both subjects (either after grade 4 or after grade 7 and that there seems to be an issue with the correctness of the unique identifiers. Thus, the present analysis estimates value-added for 1918 schools based on a full dataset for 48,529 students.

Choice of Models piloted for Estimation of Value-Added

10. The models piloted in the study are linear regression models using the scores from the national assessments in mathematics only; using scores in Bulgarian language in literature only; and using the combined total scores (the sum of the two). The linear regression model with total scores is generalized by including the random effect of the schools thus obtaining a multilevel model. The value-added estimates are measured for the total scores model in two ways: based on the regression residuals, and based on the estimated random effect. Simplified models, using the scores from Grade 4 national assessment (without other contextual characteristics) were also applied, but they do not show more significant results.

Preferred choice for a value-added model

11. Given the data available and the use of two contextual characteristics, *the type of the school* and the *language spoken at home*, it seems that the preferable choice of a model is the classical linear regression model. The multilevel model gives similar results when the starting point for the optimization procedure is

the vector of the maximum likelihood estimates from the linear model. However, a small change of the starting values gives instability of the results. The procedure is time consuming, highly dependable on the starting point, hence often failing to give result. This makes the use of the multilevel model unreliable.

Findings related to the factors affecting student assessment scores

12. Based on the data used for the analysis, it appears that the characteristics affecting the students' performance in one and the same school are:
 - the grade 4 test scores: positive influence is determined by higher input scores and higher output scores – this is true for both subjects (mathematics and Bulgarian language);
 - the gender: the survey confirms the well-known fact that girls perform better than boys for this age;
 - the language spoken at home: Bulgarian language has a positive impact on both Math and Bulgarian language scores, the Roma language has a negative impact for both exams, the Turkish language has no impact on Math scores and slightly negative impact on Bulgarian language and literature scores;
 - the type 2 of the school: the private schools have a positive impact (although they are not highly ranked) – an assumption can be made that the students from private school have a better social status, which may be a topic for further analyses;
 - the type 1 of the school: the worst effect on the response have the schools of type 3 and 4 – professional schools, sport schools or schools of arts.

Findings related to the quality of the testing instruments

13. The analysis of the test scores showed that the majority of the students received high scores on both Bulgarian language and Math national assessments at the end of Grade 4. The distributions of these parameters do not follow the normal distribution curve. However, the distributions of the same parameters for Grade 7 national assessments are characterized by much larger variance than expected. This large variance contributes to the small value of the proportion of the data explained by the model, because there is a large unexplained variance in the outcome. Obviously, the majority of the students received high scores on both national assessments; the number of low scores is very small. The distribution of the scores is not normal. A normal distribution of the results is expected for all tests that can well discriminate “lower ability” from “higher ability” students. A routine use of value-added modeling, however, would require improved testing instruments to ensure valid school value-added estimations.

Findings related to the value added measures obtained through the tested models.

14. When compared through correlation analyses, the estimated school value added and the average of the combined students score (mathematics + Bulgarian language and Literature) are characterized by a high correlation. This means that *generally speaking* the higher the test score, the higher the value-added, and vice versa, the lower the test score, the lower the value-added. This finding is in compliance with the generalized conclusions and evidence from education systems, which routinely use value-added analyses for assessment of school performance. However, at individual school levels the average and the value-added measures diverge, in some instances significantly. Despite the data limitations, inconsistencies and the limited number of contextual variables for the students, the value added measures produced by the tested models confirm the usefulness of estimating the contribution of schools to student achievements. This becomes obvious when reviewing the ranking of schools in Annex 2 to the present report. The attempt to isolate factors beyond the control of schools (in this case, the language spoken at home and the type of school) produces value added measures of performance, which diverge to a different degree from the respective ranking based on the combined average tests scores aggregated at school level.

Recommendations

15. **Complete the pilot implementation of the value added models through additional analyses.** In order to make best use of the available data, further efforts are needed to gather and analyse data in a bid to get answers to a number of questions: (i) how much are the missing data; (ii) are there enough students from each school to be included in the VAM estimates; (iii) is the percent of the students that for some reasons cannot be included in the analysis relatively small; (iv) how should be treated the students changing their schools over the period in which the value added estimates are analyzed (for example, students could be included in the value added estimate for a given school if they have been studying in this school for at least 150 days.)
16. **Re-run the models with questionnaire data for school with high response rates.** In case there is a relatively large group of schools (for example about 150 schools) with high response rate on the

questionnaire, an analysis with the existing data for this group of schools may be run to test whether the inclusion of additional contextual data make a difference.

17. **Expand the range of student-level contextual data.** This could be done either through applying questionnaire or similar survey instruments or through tapping onto the wealth of administrative data collected through the government information systems.
18. The advantage of collecting data through questionnaires is the richness, relevance and variety of information that could potentially be elicited and used for the value-added analysis. However surveys and questionnaires suffer from two significant deficiencies – low response rates and low reliability of data which depends on subjective, and not objective perceptions and may contain misleading, rather than valid information. If value-added analysis is intended to contribute to high stakes decisions, questionnaire-driven information may create controversies and compromise the value added analysis and the integrity of the high-stakes decisions. Should questionnaires be kept as a source of information, in order to avoid low response rates in future the data collection process need to be institutionalized. Further, the questionnaire needs to be simplified by reducing the number of the questions.
19. Alternatively, administrative data from different sources may be used to construct a range of variables describing the students' background characteristics. The advantage of administrative data is their reliability, routine collection and ease of use, should value added estimations be institutionalized.
20. In addition to the inclusion of a richer set of student background characteristics (through re-running the questionnaire), it can be expected that the coefficient of determination in the linear model will increase by adding new contextual characteristics – e.g. characteristics of municipalities (e.g. size, municipal revenue per capita, etc., subject to agreement and future discussion with the MEYS) or age (with a difference of at least half an year in the age of the students).
21. **Repeat the value added modeling with the next student cohort.** Including assessment data from the national assessment in 2013. The analysis can be repeated with another cohort of students – those that took Grade 4 assessment in 2010 and Grade 7 assessments that are taking place in May 2013, thus obtaining another value-added measure for each school and expanding the scope for analysis and assessment of the results obtained through the piloted value-added models to answer the principal question if the tested models produce stable value added measures over time and whether the estimated school effects are comparable over successive years for an individual school.
22. **Improve the testing instruments.** Even though the VAM are used in many educational systems to monitor school effectiveness and trigger high-stakes decisions, such use in the context of Bulgaria is premature because of the unresolved issues with data consistency and the apparent deficiencies of the testing instruments. The tests need to be constructed in a manner that enables them to capture both very low and very high performance. This is likely to affect the length of the tests and better selection of test items. A number of open questions need to be addressed: are the tests valid and reliable, are the scales sensible, how effective is supervision over the test taking process to eliminate possibilities for gaming and cheating.

Introduction

23. The objective of this report is to describe the process and summarize the results from the pilot implementation of statistical models for measuring the value-added of Bulgarian schools through analysis of the national student' assessments results. This report (as per Government request and TORs for the assignment) presents the technical aspects of the pilot and the key outcomes in terms of value-added measure for each of the schools included in the analysis. It is intended to stimulate the discussions at technical level among experts and decision makers at the MEYS, focusing on the relevance and applicability of the piloted statistical models, in the context of Bulgaria student assessment framework, and given the available data. The report documents all data processing, adjustments and procedures run as part of the school value added modeling and in this respect, it is also intended expert statisticians and researchers.
24. For long time measures of school performance in many educational systems have been based on average scores on standardized tests, there is growing recognition that there are problems in using these as measures of school performance. These measures often do not take other factors that influence educational achievement into account, such as: the native abilities of students; their socio-economic background; the influence of peers and individuals in and outside school; various events and situations that occur outside the school that might affect student learning; etc. International experience suggests that the school value-added measures (VAMs) based on a valid and reliable student assessment system are the most precise indicators for school performance developed to date. VAMs of school performance are considered accurate indicators of educational effectiveness. They are gaining in popularity and policymakers increasingly accept them as a measure of school performance. VAMs are already in use in many educational systems and widely reported in the research literature.
25. Having this in mind, it has become clear for many education stakeholders in Bulgaria that a new way of analyzing results from the national assessments is needed. Presently, the national student assessment results in Bulgaria are reported on a non-transformed raw point scale accompanied by a table that allows a transformation of the raw points to the six-grade rating scale, adopted for measuring students' performance. Based on this, each school receives an average score and some means of comparison with other schools in the country. However, datasets with the results of all schools in Bulgaria are not publicly available to avoid improper comparisons. The need for the development of fair and objective measures for school performance that could be publicly disclosed has stimulated discussions among stakeholders on the need to develop proper value-added assessment models at the country level.
26. External standardized tests were first piloted in Bulgaria in 2007. Presently national assessments are census based (covering all students in the tested cohorts), administered annually at the end of elementary (Grade 4), primary (Grade 7) and upper secondary (Grade 12, Matura) levels. The new draft law on Preschool and School Education envisions a new lower secondary education stage covering Grades 7-10, and a national, census-based, annual assessment at the end of Grade 10. However, the assessment results have not yet been put to a use that can significantly improve decision making, policy development and incentivizing school performance gains. Lingering concerns remain about the supervision of the test taking process and the arrangement to counteract gaming in test taking. In this sense, the returns to Government investments in the assessment system are lower than their potential. In order to use assessment data for improvement and accountability purposes, the Bulgarian Government needs to first address the above mentioned concerns, ensuring that test results do reflect in objective manner the knowledge of tested students. Further, the census-based assessment in Bulgaria needs to provide education stakeholders with information about the performance progress of individual schools and to facilitate the implementation of national policies for improving the quality of education.
27. Analyzing the test results through value-added analysis is a new approach for the country and the findings from this pilot are intended to inform policy makers in Bulgaria about the advantages and limitations of VAMs in the context of the Bulgarian education system, its student assessment framework and the data collected and used by the education management information system. Decomposing the gains in assessment results into school effects and effects related to student and school contextual characteristics, known as school value-added analysis, has been piloted in Bulgaria by employing both student / school level background characteristics and longitudinal tracing of assessment results of the students.

Data used for the pilot implementation of the school value-added models

28. Conducting a value-added assessment requires test scores from at least two separate points in time. As described above, in Bulgaria there are three census based national assessments. For this analysis tests scores and student level data were used for the first cohort of students that took both the national assessment test in Grade 4 (back in 2009) and the national assessment tests in Grade 7 (2012). This makes it possible to use the results of one and the same students at these two points in time in order to pilot value-added measure of the schools between grades 4 and 7. The national assessments in Grades 4 and 7 cover a number of subjects. The stable and common subject base, however, are Bulgarian language and literature and Mathematics, therefore only the test scores from these two subjects are included in the value-added analysis.
29. The linkage of the students in the merged dataset is made by their national ID number (called EGN). This number contains the information about the date of birth. Grade 4 national assessment data for the students contains, in addition to the test scores, information about the language spoken at home. A joint review by the World Bank consultants and experts of the Ministry of Education, Youth and Science (MEYS) of the available student level data and student background characteristics concluded that the value-added analysis would benefit from including more individual level information about the students. For this reason, MEYS and the Center for Control and Assessment of the Quality of the School Education (CKOKUO) prepared a questionnaire to collect additional information for all students of the cohort subject to value added analysis. The questionnaire was administered in April 2013 on the Grade 8 students – the same cohort that took Grade 7 national assessment in 2012. The questionnaire contains questions about the language spoken at home (to confirm the information collected in 2009), the members of the family, the education of the parents, the extracurricular activities of the students, and some information that helps to create socio-economic status variables (employment status of the parents, the number of the books at home, etc.). The full version of the questionnaire is included in Annex 1.
30. Unfortunately, the administration of the questionnaire took longer than expected and the response rate was very low, incurring significant delays to the pilot implementation. This triggered the decision to not use the data from the questionnaire and to pilot the models using the available data – tests scores from the national assessments of one and the same cohort of students and a limited set of student and school level characteristics (age, gender, type of school, language spoken at home). As described above, they are tested at two points of time: in 2009 when the students finished Grade 4, and in 2012 when they finished Grade 7.
31. The data provided by MEYS was structured in three files. The first file contains information about the results from Grade 4 national assessment and additional information about the student's gender and student's language spoken at home. The data were presented in table structured in the following way:
 - school ID,
 - student ID,
 - gender indicator (1 – male, 0 – female),
 - name of the student (not used in the study),
 - Bulgarian language and literature score,
 - Mathematics score,
 - “Man and Society” score (not used in the model),
 - “Man and Nature” score (not used in the model),
 - Language spoken at home (four types – three common spoken languages in Bulgaria (Bulgarian, Turkish, Roma and others).
32. The second file contains information about the results from Grade 7 national assessment. It has the following structure:
 - school ID,
 - student ID,
 - gender indicator (1 – male, 0 – female),
 - the name of the student (not used in the study),
 - Bulgarian language and literature test scores,
 - Math test scores,
 - Physics score (not used in the model),
 - Chemistry score (not used in the model),

- History score (not used in the model)
- Biology score (not used in the model),
- Geography score (not used in the model).
- Language score (not used in the model).

33. The third file contains information about the schools and some other nomenclatures.

Data processing

34. As a first step, the data from the different files were transferred in one data set – thus, a single file with the data needed for the analyses was generated through a program on Perl 5 programming language developed by the authors for the purpose of the study. The data was imported in two hash tables, one for students, with key student ID, and another for schools, with key school ID. The language spoken at home has been numerically coded as follows: Bulgarian (Code 1), Roma (Code 2), Turkish (Code 3), and other (Code 4).
35. The obtained data structures are saved as a text file with “comma separated value” format with the following structure:

1. 4-th class mathematics score,
2. 4-th class Bulgarian language and literature score,
3. 7-th class mathematics score,
4. 7-th class Bulgarian language and literature score,
5. code of the language spoken at home,
6. code of the gender,
7. indicator for school attendance in grade 4-th,
8. school type1 for grade 4 (*see Table 1*),
9. school type2 for grade 4,
10. indicator for school attendance in grade 7-th,
11. school type1 for grade 7 (*see Table 1*),
12. school type2 for grade 7.

Nomenclature codes	Type1 code
2, 6	1
3, 7	2
4, 10	3
8, 71	4
21, 32	5
50, 51, 52, 53, 54, 55, 56, 57, 58	6

Table 1. Nomenclature code of the schools

Nomenclature code	Name (explanation of the school)
1	primary (Grades 1 to 4)
2	primary and lower secondary (Grades 1 to 8)
3	school of general education (Grades 1 to 12)
4	professional school
5	primary with a nursery school
6	primary and lower secondary with a nursery school
7	school of general education with a nursery school
8	school for sports
10	lower secondary with professional classes
21	lower secondary (Grades 5 to 8)
32	Grade 5 to 12
50	school for mental derangement pupils
51	health school
52	hospital school
53	school for pupils with hearing problems
54	school for pupils with poor eyesight
55	School for pupils with speaking problems
56	social-pedagogical boarding-school
57	Instructive boarding school
58	school at prison
71	art school

Key data issues:

36. During the processing of the data the following data issues were revealed.

- Tests scores missing data: Grade 4 file contains information for 64,808 unique students, according to their student ID. After combining it with the data from Grade 7 file, there are 48,529 unique students ID, which participate in both files. This problem should be considered of significance, as it shows a certain level of data inconsistency. Nevertheless the study was performed using the result data set with the 48,529 observations for which full set of test score data are available.
- Other missing data, of less than 1% share, were identified. The obtained results should not be considered as biased because of the limited share of other missing data. The missing data are replaced by the mean of the particular variable, if it is quantitative and by its mode if it is qualitative.

37. The number of children who have participated in Grade 4 assessment is 59,288, the number of the children measured by Grade 7 national assessment is 60974. The unique students' ID are 64,808. Therefore on the basis of the data it cannot be conclude that there is a large number of drop outs. However, as it is explained above, the total number of students that are available in both datasets is 48,529. Although there are students that have taken only one exam (after grade 4 or after grade 7), this number still indicates some data inconsistency. The present analysis estimates value-added for 1918 schools using 48,529 students.

Descriptive statistics

38. There are six important characteristics at student level: language, spoken at home; gender; mathematics score from the national assessment at the end of Grades 4 and 7; Bulgarian language and literature score from the national assessment at the end of Grades 4 and 7.

Student Level Characteristics

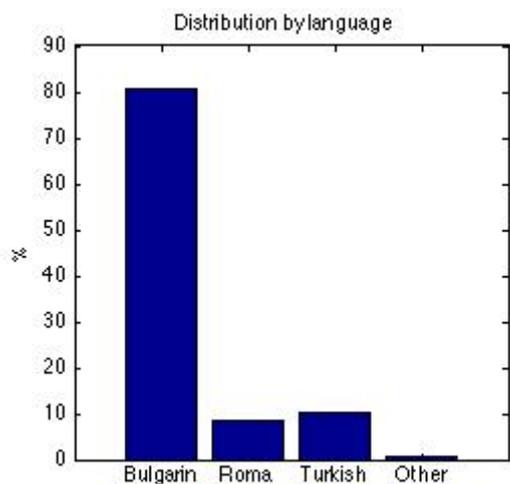


Figure 1. Distribution of language spoken at home

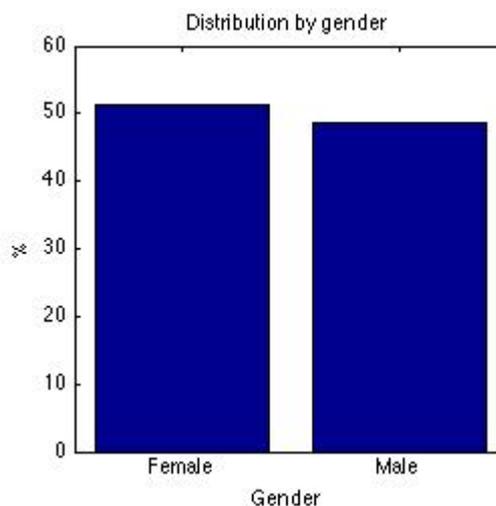


Figure 2. Distribution of students' gender

39. The distribution of the percentages of the students for different values of the language spoken at home is presented on Figure 1. Nearly 80% of the student population included in the analysis speaks Bulgarian at home, followed by Turkish, Roma and “others”. The distribution of the percentages of the students' gender is presented on Figure 2. The female students exceed by 3% the male students.

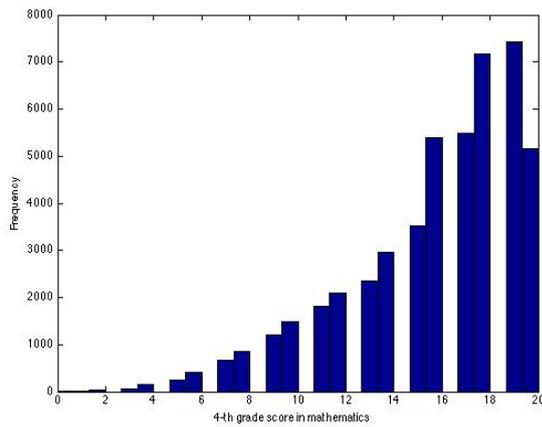


Figure 3. Histogram of grade 4 mathematics

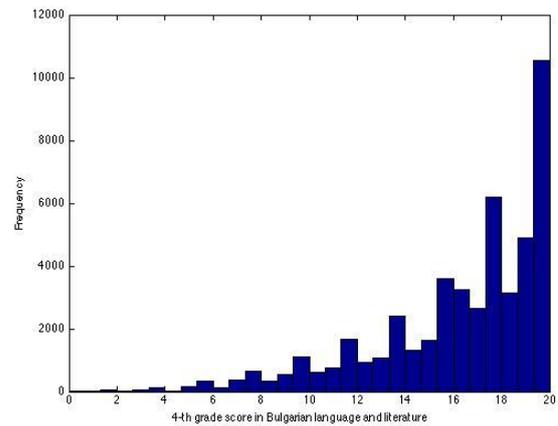


Figure 4. Histogram of grade 4 Bulgarian language and literature

40. Histograms of the scores (the graphs of the frequencies of the scores) from national assessment at the end of Grade 4 of mathematics and Bulgarian language and literature are presented on Figure 3 and Figure 4, respectively
41. Obviously, the majority of the students received high scores on both national assessments; the number of low scores is very small. The distribution of the scores is not normal (Gaussian), since its shape is not like the so-called “bell curve”. (This curve has a peak in the middle and is quickly decreasing to zero at the lowest and highest scores. The normal distribution of the results is expected for all tests that can well discriminate “lower ability” from “higher ability” students. The distributions on the figures 3 and 4 exponentially increase, indicating that the test was easy for most of the students.
42. The distributions of these parameters do not follow the normal distribution curve. The mean and the variance (the spread of the data around the mean) of the parameters are presented in Table 2. These two parameters are used as predictors in the regression models, so their distributions are not so important to the quality of the model.

Table 2. Mean and variance, Grade 4

	Mathematics	Bulgarian language and literature
Mean	15.7552	16.3228
Variance	13.0218	12.8130

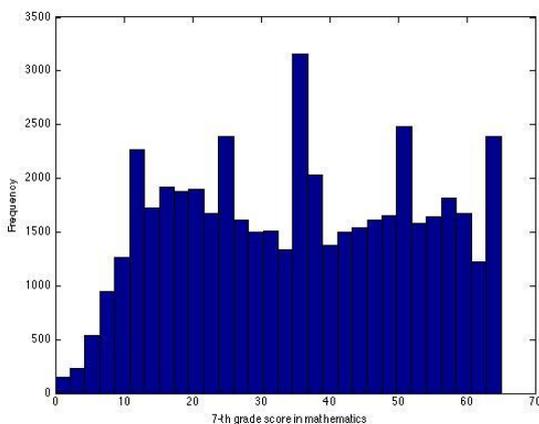


Figure 5. Histogram of grade 7 mathematics

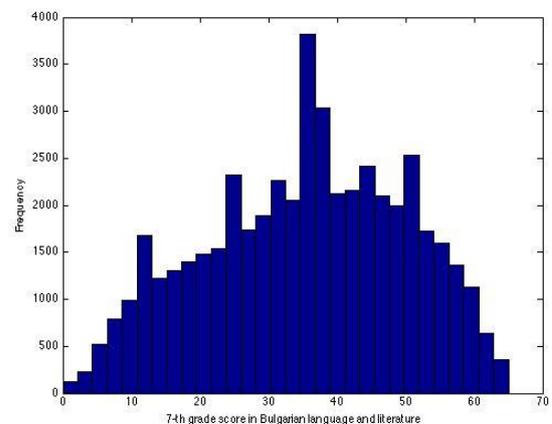


Figure 6. Histogram of grade 7 Bulgarian language and literature

43. Similarly, the distributions of the same parameters for Grade 7 national assessments are presented on the histograms on Figure 5 and Figure 6, respectively. The mean value and the variance are presented on Table 3.
44. The distribution of these two parameters is with much larger variance than expected (this is the variance if the distribution follows the normal law for data with this spread). These two parameters are response variables in the regression models in the main part of the study. As opposed to the distribution of predictors, the distribution of the responses is essential for their authentication. Interestingly, the distribution of the scores of mathematics national assessment is bimodal (has two maxima), indication inhomogeneity of the data (i.e. the data may be regarded as data from two different samples having different characteristics. This fact may be due to the participation or not-participation of all students in both modules of the assessment. This large variance contributes to the small value of the determination coefficient R^2 (the proportion of the data explained by the model) as there is a large unexplained variance in the outcome.

Table 3. Mean and variance, Grade 7

	Mathematics	Bulgarian language and literature
Mean	35.5989	35.4081
Variance	289.0382	212.3981

School Level Characteristics

45. The school level characteristics are school type (Type1 code) presented on Table 1 and the indicator whether the school is private or not. The distribution of the percentages of the schools for different types is presented on Figure 7. On Figure 8 the percentage of the private/non-private schools is presented. Obviously the distribution of these two parameters is not uniform, as there are large differences in the proportions for different values of the parameters. Most of the schools are state, primary and schools of general education.

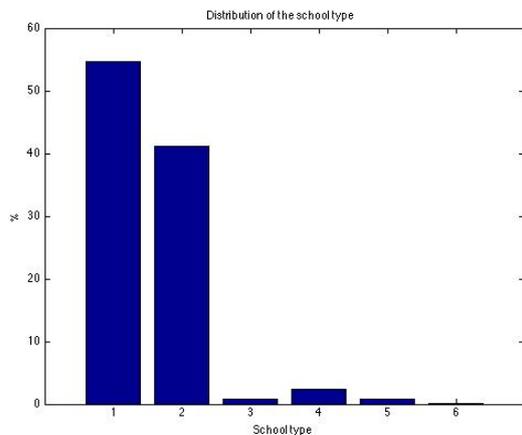


Figure 7. Distribution of school type1, grade 7

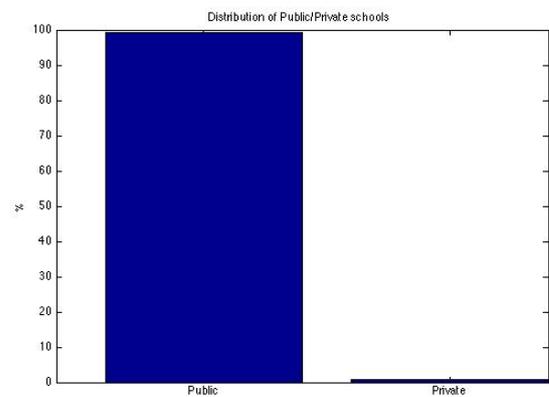


Figure 8. Distribution of school type2, grade 7

Regression based models

Linear Regression Model

46. The basic linear model, presented in this study, considers the student score at Grade 7 national assessment as a linear combination of the Grade 4 national assessment score, the student level characteristics and the school type.

$$Y_{ij2}^{mat} = \mu + \alpha Y_{ij1}^{mat} + \beta_1 X_{ij1} + \mathbf{X}_{ij2} \beta_2' + \mathbf{Z}_{i1} \gamma_1' + \gamma_2 Z_{i2} + \varepsilon_{ij} \quad (1)$$

\mathbf{X}_{ij2} is the vector of the indicators of the students' language and β_2' is a vector of the language parameters. Similarly, \mathbf{Z}_{i1} is the vector of the indicators of the school type according to Table 1 and γ_1' is

the corresponding vector of parameters. The variable X_{ij1} is the indicator for the sex of the students, and Z_{i2} is the private school indicator. The vectors are considered to be rows.

Table 4. Estimated values of the parameters for equation (1)

parameter	value	95% int-dn	95% int-up	Explanation of the parameter
μ	6.4759	4.5542	8.3975	intercept
α	2.0579	2.0200	2.0958	4 th grade score effect
β_1	1.0500	0.7951	1.3048	gender effect (1 – males)
β_2^1	4.5588	2.8331	6.2845	Bulgarian language effect
β_2^2	-3.5199	-5.2965	-1.7432	Roma language effect
β_2^3	-1.2140	-2.9780	0.5501	Turkish language effect
β_2^4	0	0	0	“others” language effect
γ_1^1	-7.3172	-7.9581	-6.6762	type1 school code effect – code 1
γ_1^2	-7.1268	-7.7726	-6.4810	type1 school code effect – code 2
γ_1^3	-12.6446	-14.3030	-10.9861	type1 school code effect – code 3
γ_1^4	-14.4907	-15.5329	-13.4485	type1 school code effect – code 4
γ_1^5	-7.2046	-8.7632	-5.6459	type1 school code effect – code 5
γ_1^6	-7.5115	-11.0925	-3.9305	type1 school code effect – code 6
γ_2	7.1163	5.5123	8.7204	type2 school code effect (1 – private)

47. The estimated values of the parameters are presented in Table 4. In this table the second column gives the estimated values of the parameters and the third and the fourth rows – the lower and the upper limits of the 95% confidence interval. All parameters can be considered as statistically significant except for β_2^3 and β_2^4 as their confidence intervals include zero. Note that the positive sign of the estimate shows a positive impact on the students’ performance, while the negative sign weakens it. The determination coefficient for this model is $R^2 = 0.2961$. It is quite low due to the thin tailed distribution of the response variable (see the histogram on Figure 5) and the large value of its standard deviation, which reflects also on the normal probability plot of the residuals, presented on Figure 9.

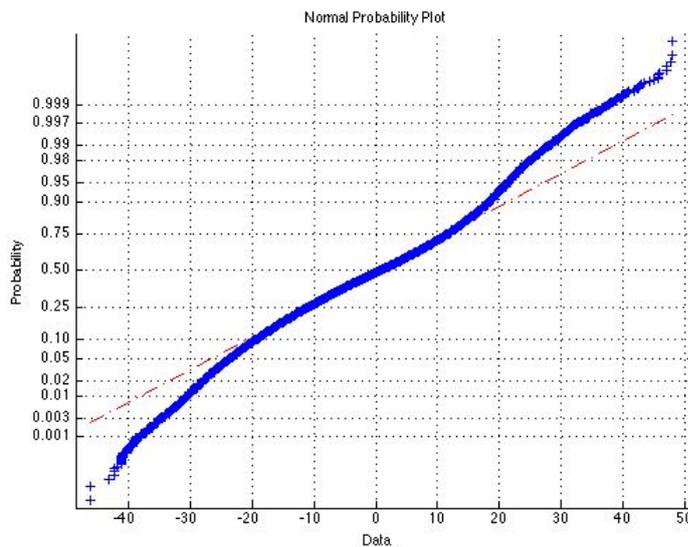


Figure 9. Normal probability distribution plot (Equation (1))

48. For normal (Gaussian) distribution, all points of the graph must lie approximately on a straight line (on Figure 9 this is the dashed line on the chart). In this case the distribution of the graph is not normal – look at the both ends. It looks like a distribution with heavy tails.

Table 5. Estimated values of the parameters for equation (2)

parameter	value	95% int-dn	95% int-up	Explanation of the parameter
μ	2.9583	1.4743	4.4423	intercept
α	2.0211	1.9906	2.0516	4 th grade score effect
β_1	3.9328	3.7350	4.1307	gender effect (1 – males)
β_2^1	3.3546	2.0248	4.6844	Bulgarian language effect
β_2^2	-5.4834	-6.8516	-4.1153	Roma language effect
β_2^3	-2.8116	-4.1701	-1.4531	Turkish language effect
β_2^4	0	0	0	“others” language effect
γ_1^1	-4.4447	-4.9381	-3.9513	type1 school code effect – code 1
γ_1^2	-4.4936	-4.9905	-3.9966	type1 school code effect – code 2
γ_1^3	-11.8883	-13.1654	-10.6112	type1 school code effect – code 3
γ_1^4	-10.0912	-10.8934	-9.2889	type1 school code effect – code 4
γ_1^5	-6.2648	-7.4652	-5.0643	type1 school code effect – code 5
γ_1^6	-1.7285	-4.4874	1.0304	type1 school code effect – code 6
γ_2	6.3130	5.0776	7.5458	type2 school code effect (1 – private)

49. The same model is applied to the Bulgarian language test score data.

$$Y_{ij2}^{bel} = \mu + \alpha Y_{ij1}^{bel} + \beta_1 X_{ij1} + \mathbf{X}_{ij2} \beta_2' + \mathbf{Z}_{i1} \gamma_1' + \gamma_2 Z_{i2} + \varepsilon_{ij} \quad (2)$$

The result presented on Table 5 shows that all parameters are statistically significant except for γ_1^6 . The determination coefficient $R^2 = 0.4319$ is larger than in the mathematics model because of the clear unimodality of the response distribution. The normal probability plot for this model is presented on Figure 10 is closer to the theoretical normal distribution.

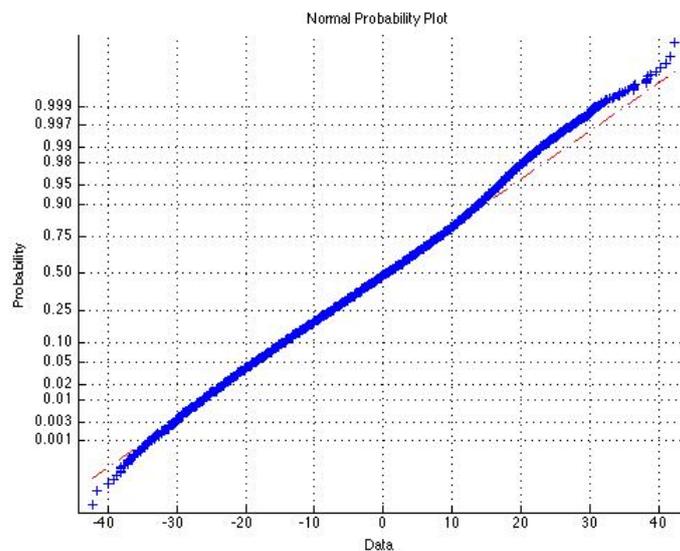


Figure 10. Normal probability plot (Equation (2))

50. The results for the total score model

$$Y_{ij2}^{tot} = \mu + \alpha Y_{ij1}^{tot} + \beta_1 X_{ij1} + \mathbf{X}_{ij2} \beta_2' + \mathbf{Z}_{i1} \gamma_1' + \gamma_2 Z_{i2} + \varepsilon_{ij} \quad (3)$$

are presented on Table 6 and Figure 11 with determination coefficient $R^2 = 0.4415$.

Table 6. Estimated values of the parameters for equation (3)

parameter	value	95% int-dn	95% int-up	Explanation of the parameter
μ	-3.1193	-6.1076	-0.1311	intercept
α	2.4434	2.4104	2.4764	4 th grade score effect
β_1	4.4311	4.0381	4.8242	gender effect (1 – males)
β_2^1	6.1654	3.5139	8.8168	Bulgarian language effect
β_2^2	-7.3737	-10.1025	-4.6448	Roma language effect
β_2^3	-3.6692	-6.3781	-0.9603	Turkish language effect
β_2^4	0	0	0	“others” language effect
γ_1^1	-10	5454	-11.5305	type1 school code effect – code 1
γ_1^2	-10.5784	-11.5704	-9.5864	type1 school code effect – code 2
γ_1^3	-22.6714	-25.2189	-20.1238	type1 school code effect – code 3
γ_1^4	-23.1396	-24.7405	-21.5388	type1 school code effect – code 4
γ_1^5	-12.8293	-15.2230	-10.4356	type1 school code effect – code 5
γ_1^6	-6.0858	-11.5870	-0.5846	type1 school code effect – code 6
γ_2	12.7191	10.2557	15.1825	type2 school code effect (1 – private)

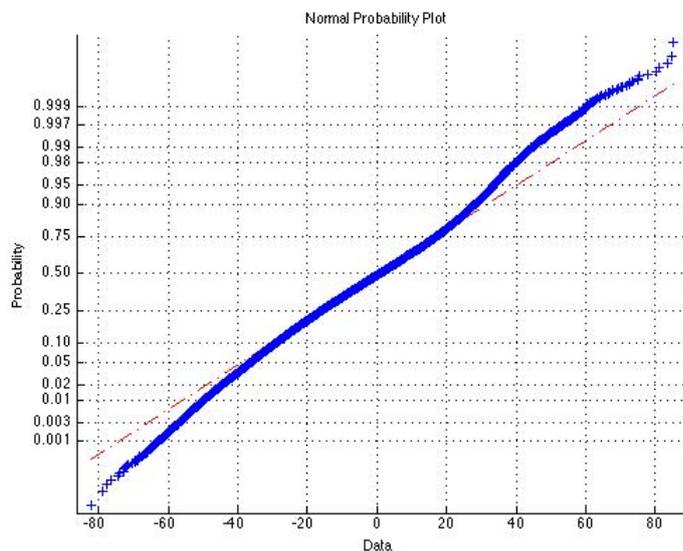


Figure 11. Normal probability plot (Equation (3))

51. The total score is the sum of the scores from mathematics and Bulgarian language and literature. Its histogram is presented on Figure 12. At a later stage, the meaning and the interpretation of the coefficients and the obtained results should be discussed with the Ministry of Education, Youth and Science (MEYS). Then the consulting group may present sensible interpretations of these statistical results.

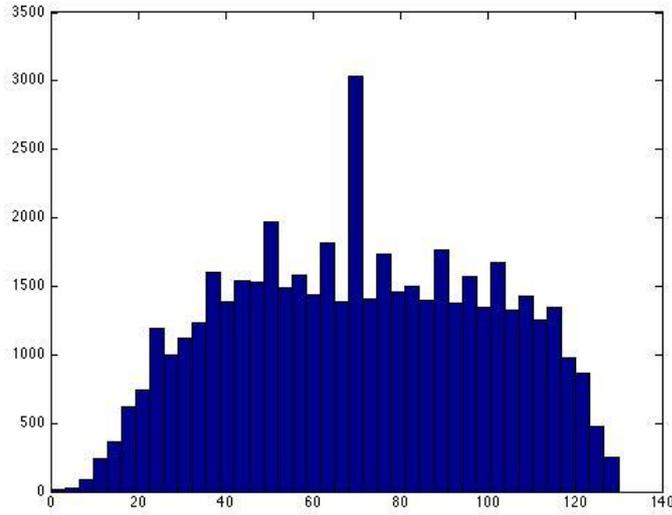


Figure 12. Histogram of the total score

Value-added Measures Based on the Regression Residuals

52. The residuals (the estimates of the error term) from the estimated total score regression are used for the calculation of the value added of the schools

$$VA_j := \frac{1}{n_j} \sum_{i \in S_{c_j}} \varepsilon_{ij} = \frac{1}{n_j} \sum_{i \in S_{c_j}} (Y_{ij2} - \hat{Y}_{ij2})$$

Here S_{c_j} is the set containing all students of the j -th school, n_j is the number of the students in this school and \hat{Y}_{ij2} is estimated predictor. Figure 13 shows the histogram of the school value added distribution. It can be seen that it follows the Gaussian law.

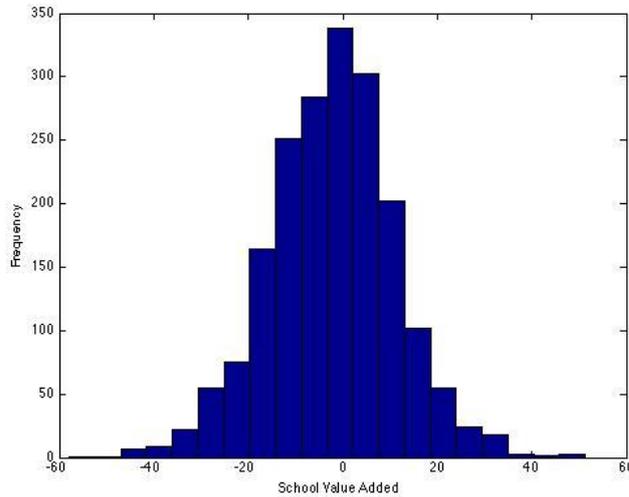


Figure 13. Distribution of the school value-added

Value-added Measures Based on the Estimated Random Effect

53. As a generalization of the last model we consider the mixed effect model (the multilevel model)

$$Y_{ij2}^{tot} = \mu + \alpha Y_{ij1}^{tot} + \beta_1 X_{ij1} + \mathbf{X}_{ij2} \beta_2' + \mathbf{Z}_{i1} \gamma_1' + \gamma_2 Z_{i2} + u_j + \varepsilon_{ij} \quad (4)$$

which includes the random effect term u_j – the random effect of the school. In fact, this is a multilevel regression with a random intercept. The estimated parameters of this model are presented on Table 7. In

this case an important role plays the variances of the random terms – the school effect and the error. They are estimated by $\sigma_u^2 = 99.8815$ and $\sigma_\varepsilon^2 = 23.5768$. The histogram of the random school effect variable is presented on Figure 14 confirming normality assumption.

Table 7. Estimation of the parameters of the model with random effect (Equation (4))

parameter	value	Explanation of the parameter
μ	-2.3582	intercept
α	2.0353	4 th grade score effect
β_1	4.2371	gender effect (1 – males)
β_2^1	5.8462	Bulgarian language effect
β_2^2	-7.3670	Roma language effect
β_2^3	-3.9434	Turkish language effect
β_2^4	0.0721	“others” language effect
γ_1^1	-9.4099	type1 school code effect – code 1
γ_1^2	-8.9013	type1 school code effect – code 2
γ_1^3	-21.1540	type1 school code effect – code 3
γ_1^4	-18.3575	type1 school code effect – code 4
γ_1^5	-11.1989	type1 school code effect – code 5
γ_1^6	-5.8928	type1 school code effect – code 6
γ_2	12.7297	type2 school code effect (1 – private)

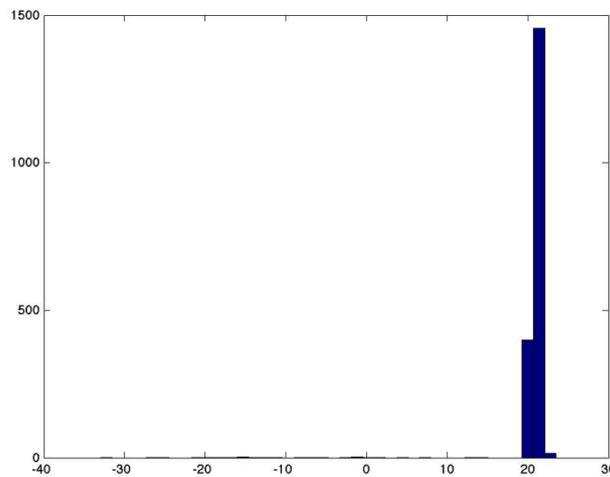


Figure 14. Distribution of the random effect

A Simplified Model

54. The simplified model

$$Y_{ij2}^{tot} = \mu + \alpha Y_{ij1}^{tot} + u_j + \varepsilon_{ij} \quad (5)$$

gives similar results presented in Table 8.

Table 8. Estimation of the parameters of the simplifies model (Equation (5))

parameter	value	Explanation of the parameter
μ	-6.0984	intercept
α	1.0854	4 th grade score effect

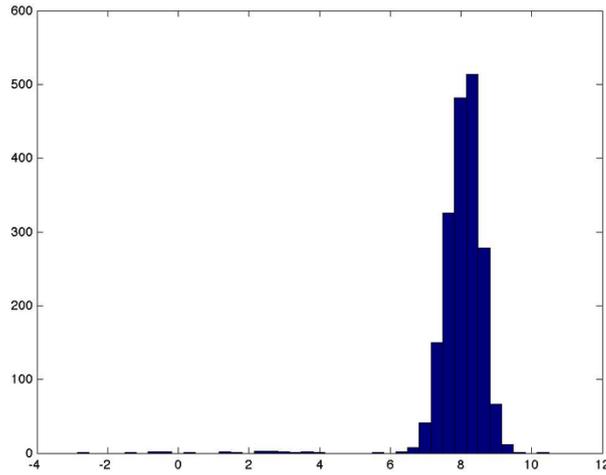


Figure 15. Distribution of the school random effect (Equation (5))

55. The histogram of the school random effect is presented on Figure 15 following again the Gaussian density function pattern. The estimated values of the variance components are $\sigma_u^2 = 85.6895$ and $\sigma_\varepsilon^2 = 23.2660$.

Shrinkage

56. Although the multilevel modeling has a number of disadvantages related to the complexity of the stochastic form and the computationally intensive procedures for its statistical estimation, it has many desirable advantages. One of the most important of them is the fact that its value added estimates "incorporate" the so called "shrinkage". Goldstein (1999) shows that the estimated school level residuals U_j are equal to the raw residuals obtained from the model's estimated fixed effects, adjusted (multiplied)

by a shrinkage factor c_j . The constant c_j is bounded by the interval $[0,1]$. Hence the residuals U_j are smaller than the raw residuals. This effect is bigger for smaller schools. As the size of school increases, the shrinkage factor gets closer to 1. This means that for large schools the value-added residuals obtained on the basis of the OLS method and of the multilevel modeling is pretty much the same, while for small schools the shrinkage factor has an important impact. More precisely, $U_j = \frac{n_j \sigma_u^2}{n_j \sigma_u^2 + \sigma_\varepsilon^2} \cdot \varepsilon_j$, where n_j is

the number of the students in the j -th school, and ε_j is the mean of raw residuals $\varepsilon_j = \frac{\sum_{i \in S_j} \varepsilon_{ij}}{n}$. The

factor multiplying the mean ε_j of the raw residuals for the j -th school is called "shrinkage factor"

$c_j = \frac{n_j \sigma_u^2}{n_j \sigma_u^2 + \sigma_\varepsilon^2}$. The degree of shrinkage depends on the size of the school. The increase of the shrinkage

for small schools can be regarded as a "compensation" for the relative lack of information (bigger standard error) for these schools, so the "best" estimate sets the predicted residual closer to the overall population value, i.e. smaller schools are "shrunk" towards the national mean.

57. At this stage, the shrinkage factor is not calculated. Based on further discussions with experts from MEYS on a sensible decision about the definition of "small schools", the shrinkage factors of the small schools will be calculated.

Impact of contextual variables represented by regression lines

58. The graph in Figure 16 represents the distribution of the average of the response (test scores after Grade 7) and the predictor variables (a linear function of the test scores after Grade 4) for a given school. Each point on the graph represents one school. The x -axis is the mean of the mathematics and Bulgarian language and literature scores at the end of Grade 4 national assessment. The y -axis is the same at the end of Grade 7 national assessment. The red (the upper) line is the expected mean test score at the end of Grade 7 national assessment for male students that speak Bulgarian language at home attending a school with $type1 = 1$ and $type2 = 0$. The green (the lower) line represents the expected mean test scores at the end of Grade 7 national assessment for students that speak Roma language under the same other conditions. The distances (in the direction of the y -axis) between the points and the corresponding lines are the school value added. The points between the red (the upper) and the green (the lower) lines represent schools that have negative impact on the score of students that speak Bulgarian language at home and positive impact on the score of students that speak Roma language at home.
59. Similar diagrams can be made to represent the impact of other contextual variables. The diagram on Figure 16 is just a sample. After discussions with experts from the Ministry of Education, Youth and Science (MEYS), the important and interesting diagrams can be presented.

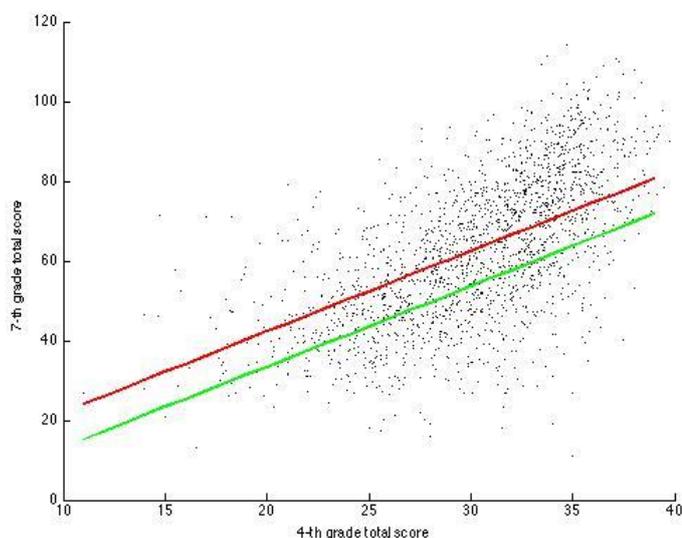


Figure. 16. Two regression lines

Correlation between the test score and the value-added

60. For the next example the test scores after Grade 7 from the schools in a Bulgarian municipality¹ are used. For these schools the estimated school value added and the average of the combined students score (mathematics + Bulgarian language and Literature) are compared through correlation analyses. The obtained correlation coefficient between the two parameters is equal to 0.8241, which shows a very good co-linearity of these two parameters. This means that (generally speaking) the higher the test score, the higher the value-added, and vice versa (the lower the test score, the lower the value-added). Figure 17 represents the average test score after Grade 7 (y -axis) versus the Grade 4 test scores (x -axis) for the schools in Bulgaria. The stars (in red) represent the schools in the municipality showing a good performance of their students above the average level in the country. Also the highest average scores after Grade 7 are from this municipality. The value-added estimates for the schools in the municipality and their ranking (compared to the ranking based on the average test scores) are presented in Annex 2.

¹ The name of the city is not stated to keep the results and data presented anonymous

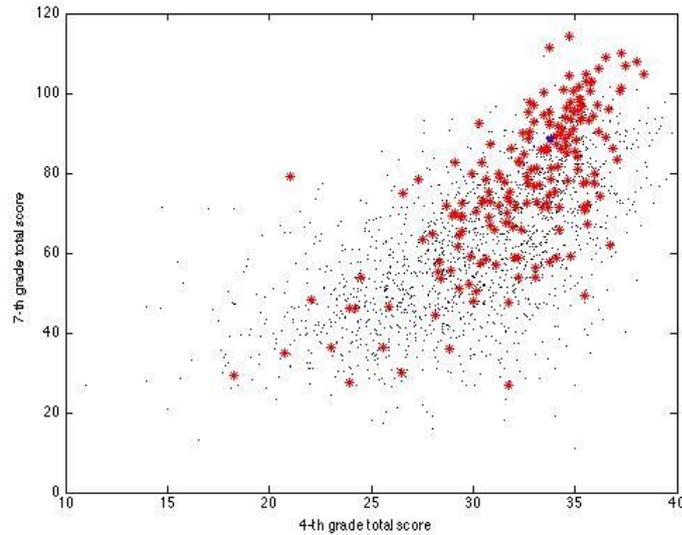


Figure 17. Value-added of schools

Key findings

61. The analysis shows that the characteristics affecting the students' performance in one and the same school are:
 - i. the grade 4 test scores: positive influence is determined by higher input scores and higher output scores – this is true for both subjects (mathematics and Bulgarian language);
 - ii. the gender: the survey confirms the well-known fact that girls perform better than boys for this age;
 - iii. the language spoken at home: Bulgarian language has a positive impact on both Math and Bulgarian language scores, the Roma language has a negative impact for both exams, the Turkish language has no impact on Math scores and slightly negative impact on Bulgarian language and literature scores;
 - iv. the type 2 of the school: the private schools have a positive impact (although they are not highly ranked) – an assumption can be made that the students from private school have a better social status – this can be further studied;
 - v. the type 1 of the school: the worst effect on the response have the schools of type 3 and 4 – professional schools, sport schools or schools of arts.
62. Despite the missing data issues and the lack of richer set of background characteristics, the results from the analysis confirm the expected characteristics that affect the response. This means that the model is adequate to the extent that it correctly identifies the effects of the variables and shows the need of improvement of the testing instruments for the national assessments. It succeeds to give results that can be used for the evaluation of the value-added of schools.
63. The first recommendation concerns the expansion of the student-level contextual data. The questionnaire developed to collect a rich set of contextual data suffered from low response rate. In order to avoid low response rate from the questionnaire in future, the data collection process need to institutionalized (as opposed to ad hoc data collection campaigns) – e.g. collection may be done by expanding the scope and range of data as part of existing, routinely administered data collection processes at school level, or tied to an event (such as taking a national assessment test) in order to ensure high response rates. It is recommended that the questionnaire becomes an integral part of the testing process by administering it as part of the national assessment instruments. This will raise the response rate and make it possible to use additional contextual data. This may well raise the response rate but there might be other issues e.g. students worried about self-identifying as Roma speaking that would skew the data or otherwise affect the mainstream testing process. Another way to improve the response rate is to simplify the questionnaire by reducing the number of the questions, leaving those that are supposed to give information strongly affected

different performance of students in one and the same school. Also, alternative approaches with pronounced advantages over questionnaires, mainly related to objectivity and reliability of data, include collection of students' background characteristics by using available administrative data across the various government sectors (e.g. employment and social assistance agency).

64. Based on the results from the analysis presented above, with the data available and used for the analysis, it seems that the preferable choice of a model is the classical linear model regression with the following two contextual characteristics: *the type of the school* and the *language spoken at home*.
65. The multilevel model gives similar results when the starting point for the optimization procedure is the vector of the maximum likelihood estimates from the linear model. However, a small change of the starting values gives instability of the results. The procedure is time consuming, highly dependable on the starting point, hence often failing to give result (i.e. to find the maximum or to find the second derivative of the inverse function). This could be explained by the non-normal distribution of the response variable (Grade 7 test score), which failed to be improved by a transformation, and its large variance, leading to an unsatisfactory appearance of the target function subject to optimization with many local maxima. The dependence on the starting point makes the use of the multilevel model unreliable. For this reason, the examples below show the value-added estimated through the linear model.

Proposed next steps and areas for further analysis

The following next steps need to be made for completing the pilot stage:

66. *Inclusion of more contextual variables.* In addition to the inclusion of a richer set of student background characteristics (through re-running the questionnaire), it can be expected that the coefficient of determination in the linear model will increase by adding new contextual characteristics – e.g. characteristics of municipalities (e.g. size, municipal revenue per capita, etc., subject to agreement from future discussion with the MEYS) or age (with a difference of at least half an year in the age of the students). All efforts need to be done to ensure high response rate to enable the use the questionnaire data, thus opening new opportunities for more precise, complete and descriptive models.
67. *Assess the impact of contextual variables from the questionnaire using a sample of schools.* In case there is a relatively large group of schools (for example about 150 schools) with high response rate on the questionnaire, an analysis with the existing data for this group of schools may be run to test whether the inclusion of additional contextual data make a difference.
68. *Including assessment data from the national assessment in 2013.* The analysis can be repeated with another cohort of students – these that took Grade 4 assessment in 2010 and Grade 7 assessments that are taking place in May 2013, thus obtaining another value-added measure for each school and expanding the scope for analysis and assessment of the results obtained through the piloted value-added models.
69. *Improving the quality of the national assessment testing instruments.* Are the measures valid, reliable, and objective? What is the quality of the test instruments? Are the scales sensible? Does not the administration of the tests allow gaming, cheating, and other shortcomings in supervision? Are the histograms of the results close to normal (Gaussian) distribution?
70. *Improving the quality and consistency of the data.* How much are the missing data? Are there enough students from each school to be included in the VAM estimates?
71. *Defining a rule for treating students changing their schools in the estimation of the value added.* Students that change the schools or leave the system. Is the percent of the students that for some reasons cannot be included in the analysis relatively small? What rule should be applied to students that change the schools? (For example: include them if they are in the school for at least 150 days.)
72. *Assessing the stability of the models.* Does the use of different VAMs substantially change the results?
73. *Assess the value added measures' consistency and validity.* Consistency through the years. Are the estimated school effects comparable over successive years for individual school?

Annex 1. Questionnaire: students background characteristics

C1 What is the spoken language at home most of the time?

(Please check only one answer).

- | | | |
|-----------|--------------------------|---|
| Bulgarian | <input type="checkbox"/> | 1 |
| Roma | <input type="checkbox"/> | 2 |
| Turkish | <input type="checkbox"/> | 3 |
| Armenian | <input type="checkbox"/> | 4 |
| Hebrew | <input type="checkbox"/> | 5 |
| Other | <input type="checkbox"/> | 6 |

C2 With whom do you live at home?

(Please check only one answer per row).

- | | | <i>Yes</i> | <i>No</i> |
|----|--|--------------------------|--------------------------|
| a) | With my mother (stepmother or foster mother) | <input type="checkbox"/> | <input type="checkbox"/> |
| b) | With my father (stepfather or foster father) | <input type="checkbox"/> | <input type="checkbox"/> |
| c) | With a brother (brothers) (including stepbrothers) | <input type="checkbox"/> | <input type="checkbox"/> |
| d) | With a sister (sisters) (including stepsisters) | <input type="checkbox"/> | <input type="checkbox"/> |
| e) | With a grandmother and/or grandfather | <input type="checkbox"/> | <input type="checkbox"/> |
| f) | With another person (e.g. cousin) | <input type="checkbox"/> | <input type="checkbox"/> |

C3 What is the occupation of your mother (or of your female guardian)?

(e.g. teacher, shop assistant, lawyer)

(If currently she does not work, please specify her last job).

C4 What is your mother's (or your female guardian's) highest education level?

(Please check only one answer).

- | | | |
|---------------------------------------|--------------------------|---|
| University | <input type="checkbox"/> | 1 |
| Secondary | <input type="checkbox"/> | 2 |
| Lower Secondary (up to grade 8 incl.) | <input type="checkbox"/> | 3 |
| Primary | <input type="checkbox"/> | 4 |
| Incomplete Primary | <input type="checkbox"/> | 5 |

C5 What did your mother do last year?

(Please check only one answer).

- | | | |
|---|--------------------------|---|
| Worked fulltime for remuneration. | <input type="checkbox"/> | 1 |
| Worked part-time for remuneration. | <input type="checkbox"/> | 2 |
| Did not work for more than a year but was looking for a job. | <input type="checkbox"/> | 3 |
| Did not work and did not look for a job (e.g. took care of our home, retired etc.). | <input type="checkbox"/> | 4 |

C6 **What is the occupation of your father (or of your male guardian)?**
(e.g. teacher, shop assistant, lawyer)
(If currently he does not work, please specify his last job).

C7 **What is your father's (or your male guardian's) highest education level?**

(Please check only one answer).

- University ₁
- Secondary ₂
- Lower Secondary (up to grade 8 incl.) ₃
- Primary ₄
- Incomplete Primary ₅

C8 **What did your father do last year?**

(Please check only one answer).

- Worked fulltime for remuneration. ₁
- Worked part-time for remuneration. ₂
- Did not work for more than a year but was looking for a job. ₃
- Did not work and did not look for a job (e.g. took care of our home, retired etc.). ₄

C9 **What was the amount of time you spent every week for studying in the form of private lessons or at tutoring/training centers or in any other type of extracurricular forms of training in the following school subjects last school year?**

*These are additional forms of training/tutoring **only in school subjects taught at school** for which you spend extra time for studying outside the regular school hours.*

(Please check only one answer per row).

	<i>Bulgarian Language and Literature</i>	<i>Math</i>	<i>Biology, Chemistry, Physics</i>	<i>Foreign Languages</i>
I spent no time for private lessons or alternative forms of extracurricular studying in these subjects	<input type="checkbox"/> ₁	<input type="checkbox"/> ₁	<input type="checkbox"/> ₁	<input type="checkbox"/> ₁
Less than 2 hours per week	<input type="checkbox"/> ₂	<input type="checkbox"/> ₂	<input type="checkbox"/> ₂	<input type="checkbox"/> ₂
2 hours or more, but less than 4 hours per week	<input type="checkbox"/> ₃	<input type="checkbox"/> ₃	<input type="checkbox"/> ₃	<input type="checkbox"/> ₃
4 hours or more but less than 6 hours per week	<input type="checkbox"/> ₄	<input type="checkbox"/> ₄	<input type="checkbox"/> ₄	<input type="checkbox"/> ₄
6 hours or more per week	<input type="checkbox"/> ₅	<input type="checkbox"/> ₅	<input type="checkbox"/> ₅	<input type="checkbox"/> ₅

C10 **Which of the following is available at your home?**

(Please check only one answer per row).

	<i>Yes</i>	<i>No</i>
a) Desk used for doing homework	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
b) Your own room	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
c) Computer you can use for school-related tasks	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
d) Educational computer programs	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
d) Internet connection	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
f) Fiction	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
g) Pieces of art (e.g. paintings)	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
h) Technical reference books	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
j) Dictionary	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
k) Smartphone	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
l) Tablet	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
m) Digital camera	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
n) DVD player	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
o) Plasma or LCD TV	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
p) Air conditioner	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂

r) Dishwasher

1

2

C11 How many books approximately are there at your home?

Normally library shelves contain around 40 books per linear meter. Do not include magazines, newspapers, comics, as well as your textbooks.

(Please check only one answer.)

0–10 books

1

11–25 books

2

26–100 books

3

101–200 books

4

201–500 books

5

More than 500 books

6

Notes and Guidelines:

Notes on the questions about the occupation of the student's parents (C3 and C6):

The professions of the mother and the father are summarized as per the categories in the National Classification of Occupations and Positions in the Republic of Bulgaria. In the answer sheets for the students, questions C3 and C6 should be formulated according to the guidelines on pages 2 and 3. The expectation is that students fill in the answer in the space provided for this. In the electronic platform the two questions will be with a drop-down menu that would contain the following categories:

- 1st category
- 2nd category
- 3rd category
- 4th category
- 5th category
- 6th category
- 7th category
- 8th category
- 9th category
- 10th category
- 11th category
- The reply of the student is not legible or is not clear.
- The student gave no answer.

Guidelines for school principals in the section describing the way in which the data about the occupational status of student's parents is to be processed (C3 and C6).

The officer inputting the data from the student's answer sheets in the electronic platform must match the profession of the mother and the father given by the student with the relevant category in the drop-down menu.

Categories are defined on the basis of the National Classification of Occupations and Positions in the Republic of Bulgaria. An eleventh category is added to the classes defined in the classification. It refers to parents described as retired, taking care of the home, unemployed, university students, deceased or other.

1st category: People holding governance/executive positions

- Members of parliament, high-level representatives of the executive (ministers, mayors, regional governors); managing and executive directors; high-level representatives of state authorities;
- High-level representatives of non-profit organizations;
- Judges (court chairpersons) and prosecutors (regional, city)

2nd category: Specialists

- Specialist in the area of scientific research in various fields of science;
- Specialists in the area of policy administration (experts in ministries, inspectorates, regional councils etc.)
- Medical specialists (physicians, nurses, midwives, veterinarian doctors, dentists etc.)
- Teachers in primary and secondary schools, university professors;
- Business and administrative specialists (financial specialists etc.)
- Lawyers, judges
- Engineers
- Librarians, journalists, interpreters/translators, archive managers;
- Musicians, actors, composers, dancers, directors, artists etc.

3rd category: Technicians and applied specialists

- Construction technicians, electric mechanics, electronics and machine mechanics, processing specialists, mining and metallurgy technicians, draftspersons, technicians in the area of monitoring of manufacturing processes, treatment plant operators, agrarian technicians;

- Technicians in the area of monitoring of manufacturing processes;
- Ship deck officers and pilots
- Applied healthcare specialists (lab specialists, dental technicians etc.)
- Business and administrative applied specialists (brokers, dealers, credit specialists, accountants, insurers etc.)
- Police inspectors, investigation officers
- Athletes and persons employed in the area of sports (coaches, instructors, athletes etc.)
- Applied specialist in the area of art, culture, and food (photographers, decorators, designers, chefs etc.)
- Technicians in the area of ICT.

4th category: Administrative support staff

- General administrative staff (secretaries, data input operators etc.)
- Administrative staff working with customers (tellers, cashiers, consultants, telephone operators, receptionists etc.)

5th category: Staff employed in the area of public services, commerce, and security

- Service staff in transport (stewards, ticket-collectors etc.)
- Guides, animators
- Cooks, waiters, bartenders, hairdressers, beauticians
- Housekeepers and domestic service staff
- Handlers and animal keepers
- Shop assistants and staff in stores, gas stations etc.
- Caregivers (child-minders, foster parents, baby-sitters, ward assistants)
- Firefighters, police officers, security officers, rescue staff

6th category: Qualified workers in agriculture, forestry, fisheries

- Gardeners and plant specialists
- Fruit growing specialists
- Animal and poultry breeders
- Beekeepers
- Forestry workers
- Fishermen and hunters.

7th category: Qualified workers and related

- Builders, masons, stonemasons, carpenters, workers doing finishing works in construction
- Plumbers
- Air conditioning mechanics
- Painters
- Metallurgy workers, founders, welders, [car] body repair workers
- Machine mechanics and repair workers
- Craftspersons and printers
- Jewelers
- Workers in the area of mounting and repairs of electric equipment
- Bakers, confectioners
- Joiners, tailors, furriers, dress cutters, shoemakers.

8th category: Machine operators and mounting workers

- Machine operators in mining (miners, quarry workers)
- Machine operators in the area of ore and mineral processing; in cement manufacturing etc.
- Machine operators of drilling equipment
- Machine operators in metallurgy and in rubber, plastic, and paper manufacturing
- Drivers of cars, busses, trains, tramlines etc. and of mobile equipment (tractor operators, digger operators etc.)
- Sailors.

9th category: Occupations requiring no special qualification

- Cleaners, washers
- Farming, forestry, and fishery workers
- Loading/offloading workers
- Street merchants
- Couriers
- Unskilled workers

10th category: Professions in the military

11th category: The parent is unemployed, retired, university student or deceased.

Annex 2. Value-added estimates for the schools in a Bulgarian municipality²

The table shows the ranking of the schools based on their combined average test score as opposed to the ranking obtained through estimation of the schools' value-added.

School ID	Ranking by Combined average test scores	Ranking by Value-added estimates	School combined average tests scores (Math + Bulgarian)	School value-added estimates (Math + Bulgarian)	Value-added estimates in a 0-100 points scale	Number of Grade 7 students	Percent of students speaking corresponded language at home				Type of school	
							Bulgarian	Roma	Turkish	Other	Type 1	Type 2
1	1	25	113.83	11.343	87	6	100	0	0	0	2	3
2	2	1	113.72	34.7	100	20	100	0	0	0	3	2
3	3	2	108.31	32.913	100	72	100	0	0	0	71	1
4	4	29	107.65	4.6799	72	17	100	0	0	0	3	3
5	5	6	107.31	24.065	98	107	100	0	0	0	2	2
6	6	5	105.56	25.223	98	73	98.63	1.37	0	0	3	2
7	7	9	104.6	22.235	97	60	98.333	1.667	0	0	2	2
8	8	26	104.17	11.092	86	15	93.333	0	0	6.667	3	3
9	9	10	102.71	21.065	96	84	100	0	0	0	3	2
10	10	8	100.91	23.167	97	11	100	0	0	0	3	2
11	11	13	100.83	19.343	96	59	100	0	0	0	3	2
12	12	11	100.55	19.93	96	86	100	0	0	0	2	2
13	13	28	100.2	10.632	85	5	100	0	0	0	3	3
14	14	7	100	23.753	98	74	100	0	0	0	3	2
15	15	14	99.937	17.888	95	157	97.452	1.274	0	1.274	3	2
16	16	24	99.917	11.815	88	6	100	0	0	0	2	3
17	17	12	99.525	19.62	96	40	100	0	0	0	3	2
18	18	17	99.525	16.994	94	103	99.029	0	0	0.971	3	2
19	19	3	98.562	30.38	99	8	100	0	0	0	71	1
20	20	15	97.618	17.366	94	72	100	0	0	0	3	2
21	21	22	97.448	14.022	91	96	100	0	0	0	3	2
22	22	18	96.89	15.915	93	95	98.947	0	1.05	0	3	2
23	23	27	96.705	10.971	86	88	100	0	0	0	2	2
24	24	19	96.296	15.302	92	54	98.148	0	0	1.852	2	2
25	25	21	95.93	14.542	91	64	98.438	0	0	1.563	3	1
26	26	34	35.532	-13.277	20	32	40.625	53.13	6.25	0	2	2
27	27	35	34.328	-13.989	18	11	100	0	0	0	2	2
28	28	31	33.364	3.2365	67	58	6.8966	93.1	0	0	2	2
29	29	33	31.389	-6.0835	36	9	22.222	66.67	0	11.11	2	2
30	30	36	31.001	-23.433	6	6	50	50	0	0	2	2
31	31	38	28.001	-35.983	0	5	20	60	20	0	2	2

² The name of the municipality is not stated specifically to ensure anonymity of the presented data. The actual school IDs are replaced by numbers. The number of included schools is different from the actual number of schools in the municipality.