Relationship between the Unified State Exam and Higher Education Performance

Tatyana Khavenson, Anna Solovyeva

Tatyana Khavenson
junior researcher at the International Laboratory for Education Policy Analysis, Institute of Education, National Research University—Higher School of Economics. Email: tkhavenson@hse.ru
Anna Solovyeva
intern researcher at the International Laboratory for Education Policy Analysis, Institute of Education, National Research University—Higher School of Economics. Email: soloveva.anna@gmail.com

Address: 13, Milyutinsky lane, Moscow, 101000, Russian Federation.

Abstract. This paper analyzes the possibility of predicting performance in higher education based on results of the Unified State Exam (USE). In particular, we test the hypothesis that USE scores in different subjects are equally efficient predictors of further academic performance. We used methods of regression analysis to assess how preliminary examinations (both composite USE scores and sores in specific subjects) affect academic performance in higher education. The research involved about 19,000 students enrolled at five Russian universities between 2009 and 2011. As long as the sample included institutions of different profiles, individual regressions were calculated for each faculty. A meta-analysis of regression coefficients was performed later to bring the data together. First-year average grades were used as the key academic performance indicator. It was found that USE scores were only related to performance in the second and the subsequent years through performance in the first year, i. e. indirectly. The research results allow to conclude that predictive capacity of composite USE scores is high enough to accept this examination as a valid applicant selection tool. The paper also analyzes the relationship between USE scores and results of subject-specific academic competitions, another student selection tool.
Keywords: Unified State Exam; high school academic competitions; preliminary examinations; prognostic validity; meta-analysis; higher education performance.

2013 was the first year when universities provided graduates who had been enrolled mostly based on their Unified State Exam scores. USE development and implementation have become the key features of the Russian education reform. As a new assessment procedure, the USE is mainly characterized by its standardized and comprehensive nature. This form of examination was designed to solve several major problems at once. First, the USE should have provided the basis for the system of school education quality assessment and at-

testation of high school graduation. Second, it should have promoted equal access to higher education for all high school graduates, regardless of their social or economic status [Bolotov, Valdman, 2012].

USE results should predict academic performance at universities: as applicants are ranked and selected by their USE scores, it is implied that students with higher scores are more talented and thus should demonstrate higher academic achievements after enrollment. Besides, the system is designed to select high school graduates based on their total USE scores in specific subjects required for admission[1]. Therefore, it is expected that further performance is predicted equally in all subjects. This paper aims to verify these two essential implications of the new examination system.

Our paramount objective is to find out how USE scores affect performance in higher education and, hence, to assess validity of the USE[2] as a preliminary examination for admissions[3]. Bearing in mind that another admission criterion is achievements in high school academic competitions, we also find it important to compare predictive validity of these selection tools in respect to further academic performance.

1. Investigating how performance in a university is predicted by preliminary examination results

As the USE was introduced not so long ago, the Russian empirical basis for research is rather small yet. Moreover, results are difficult to analyze because the existing data is poorly accessible and there are no integrated databases containing information on the USE and academic progress at universities. Nevertheless, a number of studies have already been performed on this topic [Poldin, 2011; Peresetsky, Davtyan, 2011; Derkachyev, Suvorova, 2008; Zamkov, 2012; Gordeyeva et al., 2011].

The abovementioned studies have shown that USE scores explain an average of 25–30% on the higher education progress scale, which is a rather high indicator, since academic progress is determined by a great number of factors apart from preliminary examinations. If we

---

[1] Some universities may set minimum USE scores (i. e. satisfactory thresholds), but very few of them actually use this opportunity, and the thresholds set are often rather low, which allows us to suggest that simple addition is still applied in practice.

[2] It is the prognostic value of the USE that we consider in this paper, i. e. we evaluate how accurately student progress at university may be predicted using USE results.

[3] Apart from student's 'original' skills measured by the USE, progress at university is also influenced by a great number of factors: motivation and self-control capacity [Gordeyeva et al., 2011], cultural and economic capital of the family, ethnicity, gender [Patterson, Mattern, Kobrin, 2009; Shaw et al., 2012], etc. However, we are interested in the predictive capacity of the USE in this paper, so we are not dwelling on other possible predictors of academic performance.

consider individual subjects, USE scores in mathematics and Russian turn out to be the most efficient predictors of performance. The fact that some students were enrolled as winners of academic competitions adds about 10% to explication of academic progress. However, these studies were based on narrow samples, usually including students from specific faculties within the same university, mostly in social and economic sciences; therefore, we can't use the results obtained to figure out how the USE works in selecting students for various faculties.

Standardized preliminary examinations are widely used all over the world, the most well-known ones being SAT and ACT in the US and Matura in a number of European countries. Similar procedures are also applied in Israel, Iran, Japan and China.

In the United States, every university decides which test scores, SAT or ACT, to accept for admissions. With increasing frequency, universities have been accepting scores from both, developing SAT/ACT score conversion charts. That is to say, if a high school graduate is going to enter a university, (s) he has to take either of these tests. The SAT has been applied since 1926. It has recently undergone significant changes that affected the fundamental concept of the test. The ACT was introduced in 1959, largely as an alternative to SAT. It was presented as a test to measure skills acquired at school rather than innate intellectual abilities, i. e. its results were supposed to depend mostly on student's desire and ability for learning. Eventually, however, the SAT and ACT tests converged. Both exams assess subject-specific knowledge and general study skills [Atkinson, 2009], i. e. test knowledge and skills at the same time [Zelman, 2004]. As judged by the research results, these two tests do not differ too much in their ability to predict academic progress at university [Atkinson, 2009].

The United States have accumulated a wealth of experience in conducting standardized tests and researching their validity. As compared to other preliminary examinations used worldwide, SAT and ACT are the most studied ones, with the results available to the public. That's why we refer to studies on SAT and ACT to establish some reference values for our own research.

The basic approach to analyzing SAT and ACT validity consists in assessing the linear interrelation between the test results and the indicators of academic performance by means of calculating the Pearson correlation coefficient or using regression models with academic progress indicator as a dependent variable and SAT/ACT scores as predictors. The bottom-line indicator of prediction quality is the squared correlation or determination coefficient in regression models interpreted as a part of variance of the dependent variable explained by independent variables.

Results of meta-analysis, i. e. statistical summary of results of a number of studies, have shown that the average coefficient of cor-

relation between the preliminary examination (SAT and ACT) scores and the indicator of academic progress over the whole period of studies is ranged between 0.35 and 0.46 with consideration of standard error. Thus, preliminary examinations predict 12–25% of variations in university grades (value R²) [Kuncel, Hezlett, 2007]. These are the values we are going to use as a reference point in assessing prognostic validity of the USE, since more reliable results can be obtained from summarized outputs of a number of studies than from separate studies, even if they are based on large samples.

A series of studies [Patterson, Mattern, Kobrin, 2009; Kobrin et al., 2008; Radunzel, Noble, 2012; Allen et al., 2008] demonstrate nearly the same level of SAT and ACT predictive capacity over a long period of time, which proves stability of interrelation between preliminary examination scores and success in higher education.

Along with SAT/ACT scores, some researchers also take in account the average high school diploma grades, which often turn out to be a better predictor of academic progress [Patterson, Mattern, 2012], while models considering both factors are a lot more efficient [Rothstein, 2004; Sawyer, 2010]. As long as the USE is a graduation and preliminary examination at the same time, it also serves as a high school grade.

There is ample research showing that the first year of university studies determines academic performance in all subsequent years, in final examinations [Patterson, Mattern, 2011, 2012; Radunzel, Noble 2012], and even in Master's programs [Radunzel, Noble 2012]. That's why ability to predict the first-year academic progress is an important prerequisite for prognostic validity of any test.

2. Research methodology
Step 1. Assessing the relation between USE and academic performance for each faculty

In accordance with the generally accepted methodology, we assessed prognostic validity of the USE through measuring the correlation between USE scores and further academic progress by means of linear regression analysis. We chose the first-year average grade to be the main higher education performance indicator.

Independent variables were represented by either composite USE scores (model 1) or USE scores in each specific subject (model 2):

(1) $$Y_{iy} = a + b\left(\sum_{i=1}^{k} X_{ij}\right) + e;$$

(2) $$Y_{iy} = a + b_1 X_{1i} + b_2 X_{2i} + b_j X_{ji} + e,$$

where Y_{iy} stands for performance of a student who was enrolled at a university in the year y; X_{ni} stands for the student i's USE scores in subject j; and e stands for error.

Thus, analysis of determination coefficients[4] in simple regression models (1) allows to assess prognostic validity of composite USE scores. Analysis of standardized multiple regression coefficients (2), in its turn, allows to assess USE validity for each specific subject and to compare correlation between USE scores and performance in each specific subject at university[5].

At the first stage, we regarded each faculty as an individual unit of analysis and built individual regression models for each of them. This was necessary to prevent the nature of correlation between the academic progress and the preliminary examination results from being affected by diverse assessment criteria, different variations of scores, and other factors. Bringing students from several faculties together in one model could result in underestimation of the correlation between the USE and academic performance for the selected field of study. We also built separate regression models for students enrolled in different years, as USE scales from different years are incomparable. All in all, we built around 200 models of both types—simple (1) and multiple (2)—for each faculty and student cohort at each university.

A meta-analysis of the regression coefficients obtained was performed at the second stage of research in order to summarize the results of regression models, to identify the major regularities, and to give an overall assessment of USE validity.

Meta-analysis refers to statistical methods that focus on combining results from different studies devoted to the same topic and subject. The key idea of meta-analysis is to identify the average common measure of effect size, not by just calculating the arithmetic mean of the coefficients obtained, but by weighing each of them according to their reliability. Meta-analysis is usually applied either to combine results of independent studies or to summarize results of studies with small samples, in which neither results nor statistical criteria can be regarded as reliable enough.

This paper uses a random effects model that allows to assume that coefficient effect is influenced in each specific case by a unique set of factors a researcher cannot take into consideration when combining re-

Step 2. Summarizing the results through meta-analysis

----

[4] We are going to hereinafter use interchangeably this term and the term "proportion of variance explained", or their short indication R^2.

[5] Calculation of regression models may provide two types of coefficients: standardized and non-standardized. The latter ones demonstrate how much a dependent variable (e. g. academic progress) can be increased or reduced if the independent variable (USE scores) changes by one unit. These coefficients allow to calculate the dependent variable based on the values of the independent one. As for standardized coefficients, they show the correlation between the dependent variable and each of the independent ones, thus allowing for comparison. Non-standardized coefficients are used more often to interpret results of regression analysis, but they provide too little information for our research, so we are going to use standardized coefficients in this study.

sults. This assumption is important because the relationship between a student's skills assessed by the USE at admission and her/his university grades is affected by a number of factors, such as university policies with respect to different groups of students, specific features of curricula, or methods of assessing the knowledge acquired. Besides, the sampled universities differed in size, selection criteria, number of students, and fields of study. Research on the SAT [Mattern, Patterson, 2011a, 2011b] shows that each of these factors affects the relationship between preliminary test scores and college grades. Those are the reasons why we should combine the results bearing in mind that each specific faculty has a unique set of factors affecting the extent to which a student's grades will depend on his/her skills assessed at admission.

Meta-analysis of correlation coefficient-based studies uses two modifications of random effects model building methods: the Hedges—Olkin method [Hedges, Olkin 1985] and the Hunter—Schmidt method [Hunter, Schmidt 1990]. This paper will rely upon results of the Hedges—Olkin models.

In the Hedges—Olkin method, all operations with correlation coefficients are performed after their standardization in Fisher's z-distribution. The weighted average of the coefficient is calculated as follows:

$$(3) \qquad \bar{z}_r = \frac{\sum_{i=1}^{k} Z_{r_i}\left(\frac{1}{n_i-1}+\tau^2\right)^{-1}}{\sum_{i=1}^{k}\left(\frac{1}{n_i-3}+\tau^2\right)^{-1}} \ ,$$

where $\bar{z}_r$ stands for average standardized correlation coefficient, $\bar{z}_{r_i}$ is the value of the standardized correlation coefficient in study i, n_i is the size of sample in study i, and τ^2 is the intergroup dispersion coefficient, which is calculated as follows:

$$(4) \qquad \tau^2 = \frac{Q-(k-1)}{c},$$

where k means the number of coefficients combined in meta-analysis, c is a constant used to maintain dimensionality, and Q is the coefficient of homogeneity:

$$(5) \qquad Q = \sum_{i=1}^{k}\left(n_i-3\right)\left(z_{r_i}-\bar{z}_r\right)^2.$$

Standard error of the mean is calculated as follows:

$$(6) \qquad SE\left(\bar{z}_r\right) = \sqrt{\frac{1}{\sum_{i=1}^{k}\left(\frac{1}{n_i-3}+\tau^2\right)^{-1}}} \ .$$

Table 1. **Sample description**

University	Number of sampled faculties	Number of students in the sample	Type of university	Average scores of students enrolled in 2011 based on USE results*		Number of academic competition winners among students enrolled in 2009–2011
				2010	2011	
1	14	4,653	Classical	58.2	59.9	17
2	19	6,054	Classical	59.5	59.6	129
3	23	6,618	Socio-economic	82.8	85.2	2,226
4	7	1,013	Technical	63.3	64.8	N/A
5	2	708	Technical	57.9	59.3	N/A
Total	65	19,046				

As reported by the 2011 Monitoring of Quality of Admission to Universities of Russia. Available at: http://www.hse.ru/ege/second_section/

Empirical basis

The survey sample included five universities from different regions of Russia. Four of them provided information on all of their students, while the fifth one only covered the two major faculties.

The survey used the data on students enrolled in 2009–2011. Students were selected through cluster sampling with sample units represented by universities that agreed to provide data. The overall sample included information on 65 faculties and over 19,000 students. Table 1 presents statistics on the size of the universities and the selection criteria applied by them.

3. Results
3.1. USE ability to predict long-term academic progress

Further analytical strategy is determined by the objective to establish how efficient USE scores can be in predicting long-term academic progress at university. We started with assessing the relationship between preliminary examination results and academic performance in different years of studies. To do this, we used structural equation modeling to assess direct effects models, i. e. models where USE scores had a direct effect on grades in the second and subsequent years, and indirect effects models, i. e. models where the effect was mediated by the first year of studies. The general chart of the model is shown in Figure 1. Table 2 consolidates data on the models built.

The analysis results demonstrate that USE scores predict efficiently academic progress in the first year only and have no direct effect on grades in the following years. At the same time, student performance in the second and third years depends quite a bit on the first-year grades. That is to say, the indirect relationship between USE scores and academic progress in the second and third years, mediat-

Figure 1. **Path diagram of the USE's direct and indirect effects on academic performance**
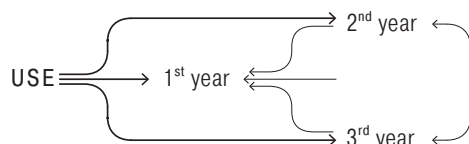

Table 2. **Results of assessing the USE's direct and indirect effects on academic performance in different years of studies**

Relationship between the variables	University					Mean value	Standard deviation
	1	2	3	4	5		
Direct effect							
USE → 1st year	0,55*	0,32*	0,48*	0,42*	0,45*	0,44	0,08
USE → 2nd year	0,13*	0,07*	−0,01	0,16*	0,10	0,09	0,07
USE → 3rd year	0,04	−0,02	0,05*	0,16*	−0,03	0,04	0,08
1st year → 2nd year	0,70*	0,71*	0,73*	0,48*	0,82*	0,69	0,13
1st year → 3rd year	0,65*	0,69*	0,66*	0,41*	0,67*	0,62	0,12
2nd year → 3rd year	0,29*	0,39*	0,33*	0,66*	0,29*	0,39	0,16
Indirect effect [a]							
USE → 1st year → 2nd year	0,39	0,23	0,35	0,20	0,37	0,31	0,09
USE → 1st year → 3rd year	0,36	0,22	0,32	0,17	0,30	0,27	0,08

The values presented are Pearson correlation coefficients. Coefficients marked with an asterisk are statistically significant at significance level 0.05.
[a] Indirect effects are calculated by multiplying the direct effect of the USE on the 1st year by that of the 1st year on the 2nd and 3rd years, respectively..

ed by the first-year grades, is rather strong: 0,3 on average for both second and third years. Thus, further analysis may only be applied to the interrelation between the USE and the first-year academic performance, while capacity of the test to predict first-year final grades will be regarded as sufficient to prove its validity.

3.2. Predictive validity of composite USE scores for different fields of study[6]

In the course of our work, we divided all sampled faculties into seven major academic fields. Further generalization by means of meta-analysis procedures was performed for each academic field in-

---

[6] At the initial stages of research, we discovered that the USE's predictive validity hadn't been increased or reduced in any of the universities from our

Table 3. **Number of models combined in meta-analysis**

Academic fields	Faculties	Number of models
Mathematics and Computer Science	Mathematics and Computer Science	26
Physics and Engineering	Physics Engineering	56
Natural Sciences and Medicine	Biology Ecology Geography Chemistry Geology Medicine	32
Economics	Economics	21
Management, Marketing, Sociology, Public Relations	Management, Marketing, Sociology Customer Services, Advertizing, Public Relations	33
Philology and Journalism	Philology Journalism	18
Humanities	History Philosophy, Culturology, Oriental Studies, Political Science	18

dividually. Table 3 presents these academic fields of study and specifies the number of basic regression models (1) and (2) combined in meta-analysis. In combining these models, we were guided not only by generally accepted classifiers but also by completion of groups in our data.

Figure 2 shows what proportion of variance[7] in first-year grades is explained by results of preliminary examinations within each academic field[8].

R^2 values in the range 0.13–0.3 mean that 13–30% of first-year grades are explained by the composite USE scores the student was enrolled with. This is pretty much, given that we didn't consider any other indicators that could possibly influence academic performance. The results obtained almost fit into the reference values 0.12–0.25 that we established after reviewing the studies on SAT and ACT prognostic validity.
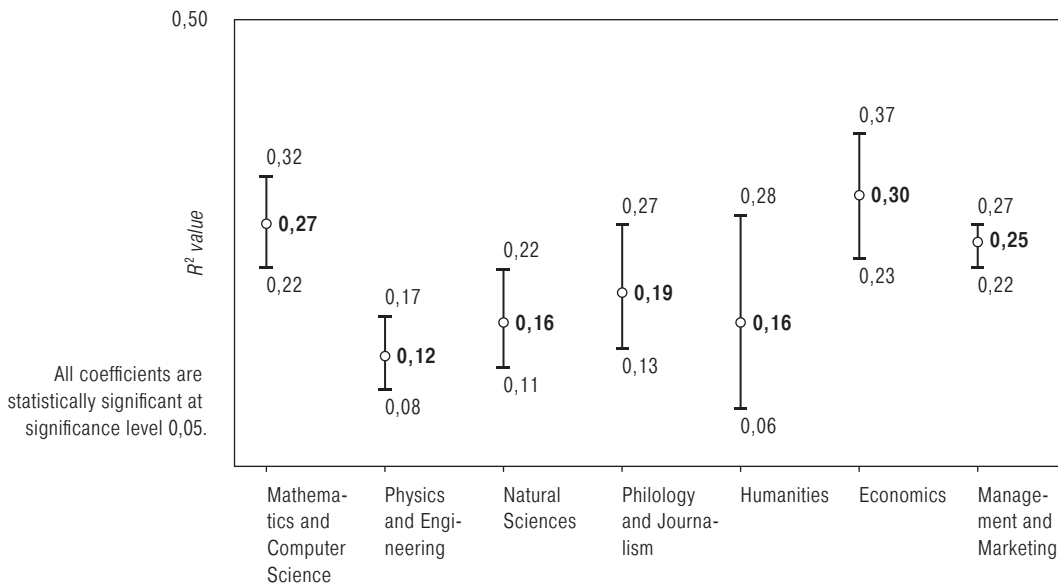
---

sample in 2009–2011; therefore, we didn't analyze the test's predictive capacity for different years.

[7] Determination coefficient R^2 of regression models of type one (1). Lines hereinafter indicate the confidence interval.

[8] Mean values were calculated through meta-analysis procedures.

Figure 2. R^2 **mean values for different academic fields**

All coefficients are statistically significant at significance level 0,05.

As confidence intervals for PVEs of most academic fields overlap, we can conclude that the USE predicts academic progress with nearly the same efficiency for each field of study. There are some differences, though: composite USE scores are better predictors of grades in faculties of Economics, Mathematics and Computer Science, Management and Marketing, as compared to Physics and Engineering, where R^2 values are statistically way lower than those in the three abovementioned domains. Overall, however, we can claim that composite USE scores have a rather high predictive validity in all academic fields.
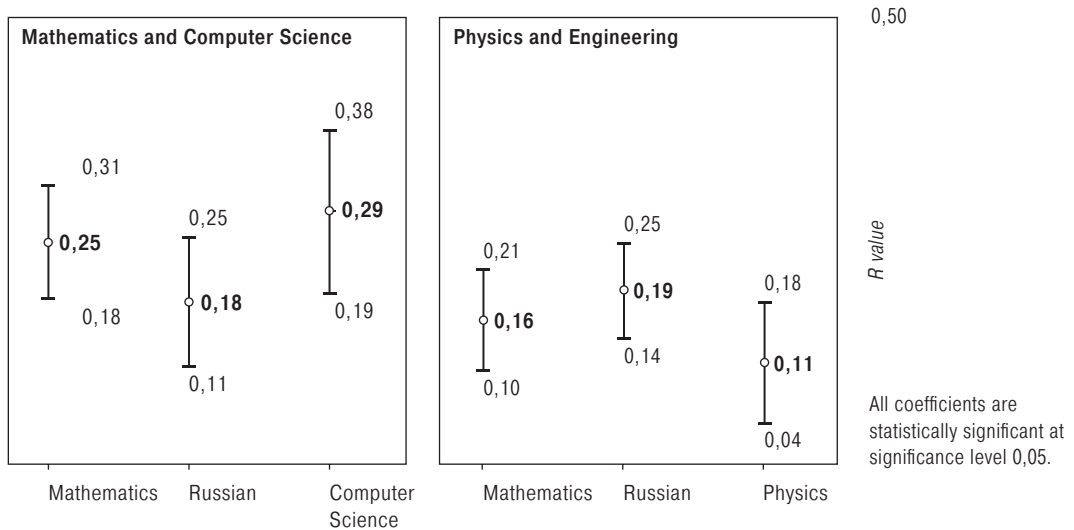
The reasons for differences in predictive capacity of composite USE scores become clear when we investigate the composition of predictive validity of different subjects included in composite USE scores. For this purpose, we assessed models (2) and summarized their results using meta-analysis. Figures 3–6 show differences in standardized regression coefficients[9] for USE scores in different subjects across various academic fields.

In faculties of mathematics and computer science, regression coefficients were statistically significant for USE scores in all subjects, major subjects (mathematics and computer science) being the most important predicators of further academic progress. This means that university grades were determined by USE scores in all subjects, and most of all in major ones, i. e. mathematics and computer science.

_____

[9] These coefficients may be interpreted as the relationship between USE scores in specific subject and academic performance at university.

Figure 3. **Mean values of standardized regression coefficients in spe-
cific subjects for faculties of Mathematics and Computer Science
and those of Physics and Engineering**

The coefficient for USE scores in Russian was nearly as high as those in major subjects. Conversely, USE scores in Russian and mathematics appeared to be better predictors than those in the major subject in the Physics and Engineering field. It is interesting that all regression coefficients were lower in this domain than in Mathematics and Computer Science.

Interpreting regression coefficients for USE scores in different subjects within Natural Sciences is rather difficult, as we combined most diverse faculties in this field[10], differing in their sets of USE subjects required for admission. That's why we can't say whether USE scores in mathematics and Russian are better predictors of academic performance than major subjects in natural sciences. However, the existing results help estimate predictive validity of the USE for each subject individually.

USE scores in mathematics were the best predictor of grades in faculties of Economics, Marketing and Management; scores in Russian also had a rather high coefficient. Coefficients of foreign language and social theory were pretty much the same, way lower than those of Russian and mathematics.

In faculties of philology and journalism, the highest predictive capacity was demonstrated by USE scores in history, their coefficient being much higher than those of Russian and literature, which are major subjects. For the rest of humanities (philosophy, culturology,

---

[10] This was done to avoid sample size restrictions in meta-analysis.

Figure 4. **Mean values of standardized regression coefficients in specific subjects for faculties of Natural Sciences**
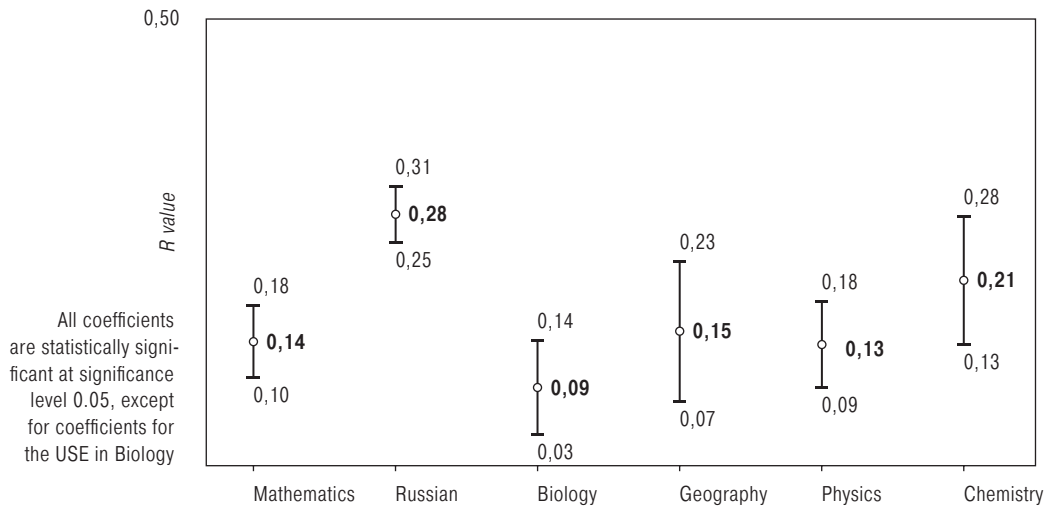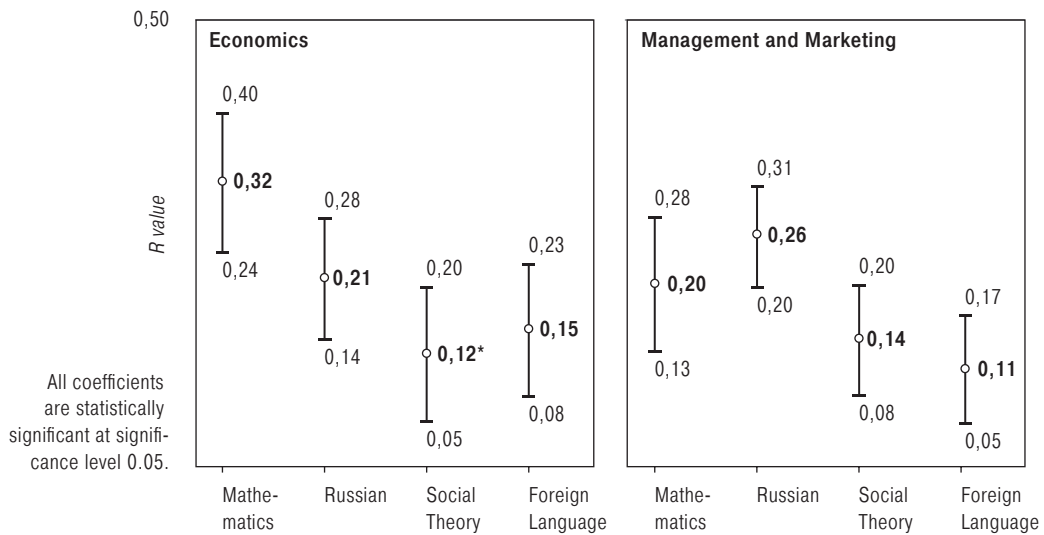

All coefficients are statistically significant at significance level 0.05, except for coefficients for the USE in Biology

Figure 5. **Mean values of standardized regression coefficients in specific subjects for faculties of Economics, Marketing and Management**

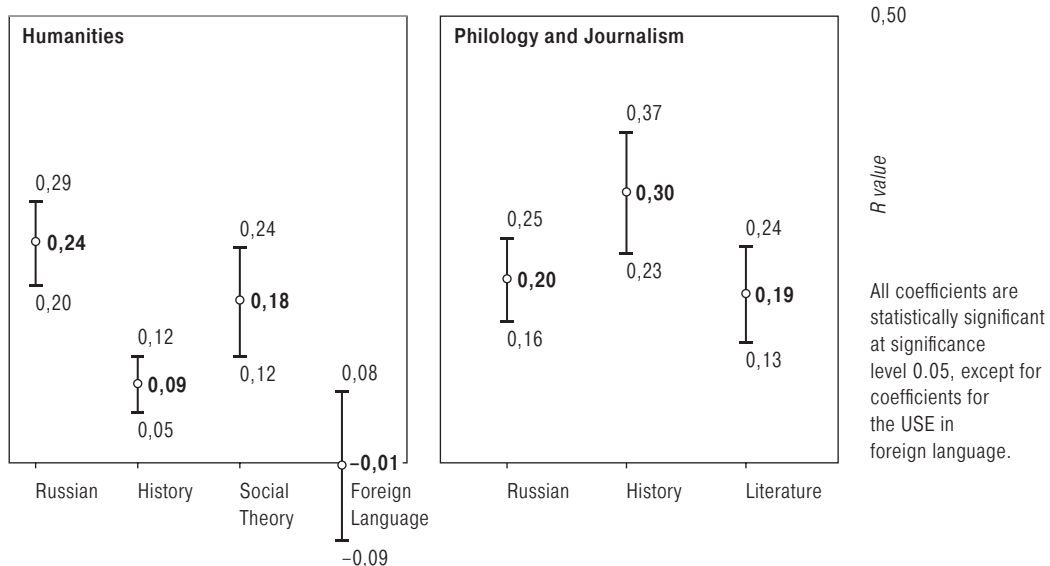All coefficients are statistically significant at significance level 0.05.

oriental studies, political science, and history), only USE scores in social theory and Russian were meaningful, while history and foreign language didn't play any prominent role.

As soon as the extent to which USE scores in different subjects affect performance in higher education is interpreted as assessment of their predictive validity, the latter can be ranged (Table 4).

Figure 6. **Mean values of standardized regression coefficients in specific subjects for faculties of Humanities, Philology and Journalism**

All coefficients are statistically significant at significance level 0.05, except for coefficients for the USE in foreign language.

Overall, the analysis of predictive validity of the USE in specific subjects has shown that USE scores in mathematics and Russian appear to be better predictors of grades than scores in major subjects, which is true for almost all academic fields. Sometimes USE scores in major subjects even turn out to be the poorest predictors, showing less effect than scores in Russian, mathematics and foreign languages.

We can suggest a number of reasons for the highest validity level of USE scores in mathematics and Russian.

First, these USE subjects have the most clearly expressed double function: testing the knowledge acquired at school and assessing the competencies required for admission to university. The USE tests in Russian and mathematics are taken both by those who only need a high school diploma and those willing to use their test scores to enter a university, including field-specific faculties. Probably, the very content of the test in these two subjects is more differentiated, increasing precision in evaluating students, i. e. differences between applicants in their mathematics and Russian USE scores reflect differences in their competencies more precisely than results of the rest of the subjects[11].

---

[11] We proceed here from the assumption, unproven but reasonable to us, that optional USE subjects are chosen by more motivated students, either because they like the subject and feel confident in it or because it is required by the university they apply to, or for any other reason. Anyway, optional subjects are definitely rarely taken by those who have had low grades in these subjects at school.

Table 4. **Average predictive validity of the USE
in different subjects across all faculties**
(Standard errors in parentheses)

USE subject	Mean value of standardized regression coefficient[2]
Computer science	0,29 (0,05)
Russian	0,22 (0,03)
Mathematics	0,21 (0,04)
History	0,19 (0,05)
Chemistry	0,21 (0,07)
Literature	0,19 (0,05)
Physics	0,16 (0,04)
Social theory	0,15 (0,04)
Foreign language	0,13 (0,04)
Geography	0,15 (0,08)
Biology	0,09 (0,05)

Besides, as it has been shown above, the US studies on SAT and ACT validity demonstrate that academic performance at university can be predicted more efficiently if the average high school grade is considered. In this case, we can say that the USE serves as the high school final grade to some extent.

Second, mathematics and Russian are compulsory USE subjects, which makes a center of attention for USE developers and everyone engaged in preparation for the test. Schools apply most of their efforts to compulsory USE subjects, thus allowing students to prepare for them on their own, without resorting to coaches or extra-curricular courses. Thus, students from different social classes have equal chances of passing the tests, and non-intellectual factors exert the least influence in these subjects.

Third, this phenomenon may be explained not through subject-specific features of the test, but through first-year university curricula. In virtually all fields of study, first-year students have a great number of interdisciplinary subjects that do not reflect specifics of the field. That is why USE scores in Russian and mathematics, which are also interdisciplinary subjects, demonstrate the strongest relationship with higher education grades.

Fourth, finally, we may suggest that Russian and mathematics are simply fundamental high school curriculum subjects, and knowledge of them is associated with basic competencies required for successful learning in any field.

Conversely, we absolutely didn't expect that USE scores in field-specific subjects would often be the poorest predictors of academic progress. We believe that this has to do with the test content in these subjects and with teaching techniques applied at universities. High performance in field-specific subjects at university is provided rather by ability to understand and to learn something new than by willingness to cram up rules and facts. We can thus suggest that USE scores in field-specific subjects measure rather factual knowledge of graduates, i. e. the skills they acquired at school, but not overall understanding of the subject or ability to research in the specific field. In other words, the USE tests knowledge, not capabilities here[12].

High school academic competitions are regarded as a form of dealing with gifted children who perform the best at school. Winners of a number of competitions in Russia are entitled to get enrolled at a university on a non-competitive basis. Thus, the USE and academic competitions are complementary forms of selection: competitions reveal the most talented students, while the rest of the applicants are selected through the USE test. Evaluating prognostic validity of academic competitions as selection tools and comparing it with effectiveness of the USE selection power was one of our primary goals.

Our sample included only one university with a really large number of students enrolled subsequent to competition results, that's why we only used data on that university to compare predictive capacity of the USE and academic competitions. Winners of contests are not required to provide any USE scores to be admitted to a university. For this reason, approximately 20% of competition winners at each faculty hadn't passed the USE in at least one qualifying subject.

First of all, we wanted to see if students enrolled as competition winners had higher USE scores than others. For each faculty, we calculated the difference between composite USE scores of students admitted based on USE results and those admitted as winners of academic contests.

Figure 7 shows mean values across three groups of faculties: 1) faculties accepting results of mathematical competitions; 2) faculties accepting results of competitions in mathematics and economics; and 3) faculties accepting results of competitions in any other subjects. The average composite USE scores were nearly the same for competition winners and other applicants in almost all faculties.
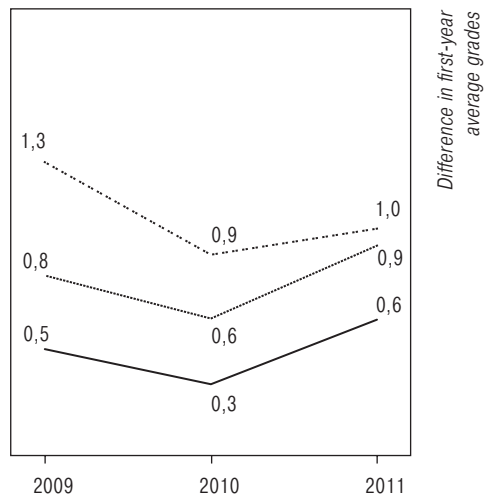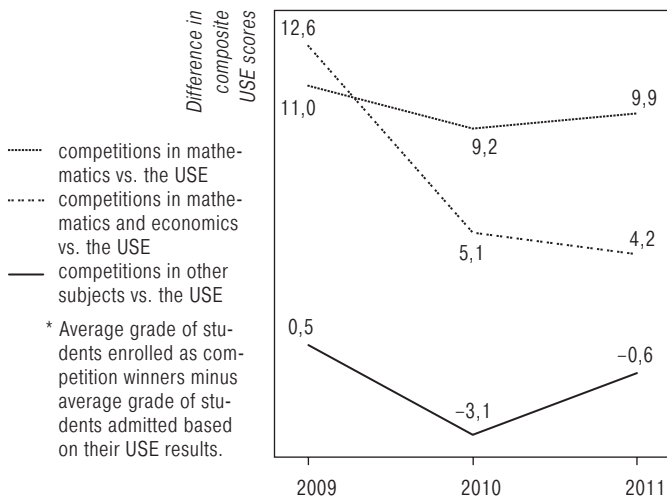
3.3. Comparing USE scores with results of high school academic competitions

---

[12] These assumptions could be verified if we had studied the relationship between USE scores in field-specific subjects and academic progress in all of the following years. However, it was impossible to do using the data that we had at our disposal. Besides, the vector chart of the relationship between composite USE scores and grades in different years of studies (see above) demonstrates clearly that USE scores affect mostly performance in the first year, which, in its turn, determines further progress.

Figure 7. **Difference in composite USE scores between students enrolled as competition winners and students admitted based on their USE results**

Figure 8. **Differences in first-year average grades between students enrolled as competition winners and students admitted based on their USE results***

Difference in composite USE scores

.... competitions in mathematics vs. the USE

- - - - competitions in mathematics and economics vs. the USE

——— competitions in other subjects vs. the USE

* Average grade of students enrolled as competition winners minus average grade of students admitted based on their USE results.

Difference in first-year average grades

12,6
11,0
9,2
9,9
5,1
4,2
0,5
−0,6
−3,1

2009   2010   2011

1,3
1,0
0,9
0,8
0,9
0,5
0,6
0,6
0,3

2009   2010   2011

The exception was the group of faculties where most students were enrolled as winners of competitions in mathematics and mathematical economics: such students had USE scores that were statistically much higher than those of students admitted based on their test results. Therefore, mathematical competitions differentiate students similarly to composite USE scores: applicants who succeeded in competitions generally obtain higher USE scores than others, which means that results of mathematical competitions correlate with USE scores.

Next, we analyzed if there was any difference in grades between students enrolled as competition winners and the rest of the students. The mean difference in first-year grades was rather small, with statistically significant contrasts in about 50% of faculties. Figure 8 shows differences in the average grades between students enrolled as competition winners and students admitted based on their USE results, across all of the three groups mentioned above. Just as with the USE, competitions in mathematics and economics predicted academic performance better than competitions in other subjects.

Thus, students enrolled as competition winners only have much higher USE scores and academic grades in some of the faculties, mostly associated with mathematics and economics, while achievements in other faculties appear to be pretty much the same for both competition winners and students admitted based on their USE scores.

4. Conclusion

The main research objective of this paper was to assess the USE as an applicant selection tool. For that purpose, we tested its ability to predict further academic progress of students in different fields of study. We measured validity of both composite USE scores—the basic selection tool—and USE scores in specific subjects accounting for the cumulative total. We also analyzed specific aspects of interaction between the USE and another applicant selection tool, academic competitions, regarding them as complementary techniques.

Results of the analysis reveal that predictive capacity of composite USE scores is sufficient to recognize the exam as a valid applicant selection tool. In our measurements, we were guided by validity indicators of SAT and ACT standardized tests implemented in the United States, which may serve reliable reference values, given the extensive experience of related research and development. The mean value of determination coefficient for composite USE scores models is 0.20, which means that approximately 20% of grades of first-year[13] students in different academic fields are only explained through their scores in preliminary examinations, i. e. the USE. The coefficient varies from 15 to 35% across faculties. In this context, we believe that multifunctionality of the USE as a combination of high school final grade and preliminary examination plays a rather positive role, increasing the test's ability to differentiate applicants. Moreover, composite USE scores reflect not only the level of preparedness for higher education, but also the level of knowledge acquired at school.

Predictive capacity of USE scores in specific subjects accounting for composite USE scores is relatively the same, but scores in mathematics and Russian are better predictors of grades in almost all of the academic fields. Conversely, USE scores in field-specific subjects often appear to be poorly related to performance at university.

Having compared groups of students admitted based on their USE results and those enrolled as competition winners, we found that differences, both in average USE scores and in first-year average grades, only took place in faculties accepting mostly results from competitions in mathematical economics. However, indicators of students enrolled as competition winners were only a little higher in both cases. Results that we have obtained do not absolutely disprove the idea that competitions are won by gifted children only, but nei-

---

[13] It is the first year of studies that determines academic performance in all subsequent years, so grades in the second and the following years are largely predicted by first-year final grades, not by preliminary examination results. We found it was true for all the universities included in the sample. Therefore, USE scores are indirectly related to long-term academic progress at university, but direct relationship between preliminary examinations and first-year grades, which account for subsequent performance, is quite enough to prove the USE validity.

ther do they prove that students of this category perform much better at university than those who were admitted based on USE scores.

Thus, for the sampled universities, the USE has proved to be a valid applicant selection tool, which allows, along with academic competitions, to identify the most talented applicants and to predict their success in higher education. Similar research should be conducted on larger samples with more universities and, vice versa, on smaller samples with a deeper analysis and a broader focus of research. For instance, we could assess the relationship between USE scores in field-specific subjects and grades in these subjects in the first and subsequent years of studies, which would allow to test a number of hypotheses we made in paragraph 3.2. Besides, it is necessary to measure the impact of factors that can build the indirect relationship between preliminary examination results and long-term academic progress. This paper has only touched upon first-year grades as such factor, but socioeconomic status of a student or her/his family may play a role, too.

References

1. Allen J., Robbins S., Casillas A., Oh I.—S. (2008) Third-Year College Retention and Transfer: Effects of Academic Performance, Motivation, and Social Connectedness. *Research in Higher Education*, vol. 49, pp. 647–664.
2. Atkinson R. C. (2009) *The New SAT: A Test at War with Itself.* Available at: http://www.rca.ucsd.edu/speeches/AERA_041509_Speech_Reflections_on_a_Century_of_College_Admissions_Tests.pdf (accessed 31 January 2014).
3. Bolotov V., Valdman I. (2012) Kak obespechit effektivnoye ispolzovaniye rezultatov otsenki obrazovatelnykh dostizhenii shkolnikov [How to Provide Efficient Use of Academic Performance Assessment Results]. *Obrazovatelnaya politika*, no 1 (57), pp. 36–42.
4. Derkachyev P., Suvorova I. (2008) Yediny gosudarstvenny ekzamen kak sposob otsenki potentsiala k polucheniyu vysshego obrazovaniya [The Unified State Exam as a Way of Assessing Potential for Success in Higher Education]. *Sbornik statey aspirantov HSE* [Collection of articles by PhD students from the Higher School of Economics], Moscow: HSE, pp. 34–64.
5. Field A. P. (1999) A Bluffer's Guide to Meta-Analysis I: Correlations. *Newsletter of the Mathematical, Statistical and Computing Section of the British Psychological Society*, vol. 7, no 1, pp. 16–25.
6. Field A. P. (2001) Meta-Analysis of Correlation Coefficients: A Monte Carlo Comparison of Fixed-and-Random-Effects Methods. *Psychological Methods*, no 6, pp. 161–180.
7. Gavaghan D. J., Moore R. A., McQuay H.J. (2000) An Evaluation of Homogeneity Tests in Meta-Analyses in Pain Using Simulations of Individual Patient Data. *Pain,* vol. 85, no 3, pp. 415–424.
8. Gordeyeva T., Osin Y., Kuzmenko N., Leontyev D., Ryzhova O., Demidova Y. (2011) Ob effektivnosti dvukh system zachisleniya abituriyentov v khimicheskiye vuzy: dalneyshii analiz problem [On Efficiency of the Two Admission Systems in Chemical Universities: A Further Analysis]. *Yestestvennonauchnoye obrazovaniye: tendentsii razvitiya v Rossii i mire (red. V. Lunin)* [Scientific Education: Developmental Trends in Russia and Worldwide (ed. V. Lunin)], Moscow: MGU, pp. 88–100.

9. Hedges L. V., Olkin I. (1985) *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.
10. Hunter J. E., Schmidt F. I. (2004) *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings.* London: Sage.
11. Hunter J. E., Schmidt F. L. (1990) *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Newbury Park, CA: Sage.
12. Kobrin J. L., Patterson B. F., Shaw E. J., Mattern K. D., Barbuti S. M. (2008) *Validity of the SAT for Predicting First-Year College Grade Point Average.* Research Report No. 2008–5. Available at: http://research.collegeboard.org/publications/content/2012/05/validity-sat-predicting-first-year-college-grade-point-average (accessed 31 January 2014).
13. Kuncel N. R., Hezlett S. A. (2010) Fact and Fiction in Cognitive Ability Testing for Admissions and Hiring Decisions. *Current Directions in Psychological Science*, vol. 19, no 6, pp. 339–345.
14. Kuncel N. R., Hezlett S. A. (2007) Standardized Tests Predict Graduate Students Success. *Science*, vol. 315, no 5815, pp. 1080–1081.
15. Mattern K. D., Patterson B. F. (2011a) *The Relationship between SAT Scores and Retention to the Third Year: 2006 SAT Validity Sample.* College Board Statistical Report No 2011–2.
16. Mattern K. D., Patterson B. F. (2001b) *The Relationship Between SAT Scores and Retention to the Fourth Year: 2006 SAT Validity Sample.* College Board Statistical Report No. 2011–6.
17. Mattern K. D., Patterson B. F., Shaw E. J., Kobrin J. L., Barbuti S. M. (2008) *Differential Validity and Prediction of the SAT.* College Board Research Report No 2008–4.
18. National Research Council (1992) *Combining Information: Statistical Issues and Opportunities for Research*. Washington, D.C.: National Academy Press.
19. Patterson B. F., Mattern K. D. (2012) *Validity of the SAT for Predicting First-Year Grades: 2009 SAT Validity Sample.* College Board Statistical Report No 2012–2.
20. Patterson B. F., Mattern K. D., Kobrin J. L. (2009) *Validity of the SAT for Predicting FYGPA: 2007 SAT Validity Sample.* College Board Statistical Report No 2009–1.
21. Patterson B. F., Mattern K. D. (2011) *Validity of the SAT for Predicting Forth-Year Grades: 2006 SAT Validity Sample*. College Board Statistical Report No 2011–7.
22. Peresetsky A., Davtyan M. (2011) Effektivnost EGE i olimpiad kak instrumenta otbora abituriyentov [Efficiency of the USE and Academic Competitions as Student Selection Tools]. *Prikladnaya ekonometrika*, no 3, pp. 41–56.
23. Poldin O. Prognozirovaniye uspevayemosti v vuze po rezultatam EGE [Predicting Higher Education Performance by USE Points]. *Prikladnaya ekonometrika,* no 1, pp, 56–69.
24. Radunzel J., Noble J. (2012) *Predicting Long-Term College Success through Degree Completion Using ACT Composite Score, ACT Benchmarks, and High School Grade Point Average*. ACT Research Report Series. No 2012–5. Iova City: ACT Inc.
25. Rothstein J. M. (2004) College Performance Predictions and the SAT. *Journal of Econometrics*, no 121, pp. 297–317.
26. Sawyer R. (2010) *Usefulness of High School Average and ACT Scores in Making College Admission Decisions*. ACT Research Report Series No 2010–2.

27. Shaw E. J., Kobrin J. L., Patterson B. F., Mattern K. D. (2012) *The Validity of the SAT for Predicting Cumulative Grade Point Average by College*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA. 10 April, 2011.
28. Zamkov O. (2012) Otsenki EGE kak indicator posleduyushchikh akademicheskikh uspekhov studentov mezhdunarodnoy programmy po ekonomike [USE Points as an Indicator of Subsequent Academic Success in the International Economics Program]. Proceedings of the *13th International Scientific Conference on Economy and Society Development (Moscow, Russia, May 3–5, 2012)* (ed. Y. Yasin), Moscow: HSE, pp. 304–313.
29. Zelman M. (2004) Osobennosti EGE v kontekste opyta obrazovatelnogo testirovaniya v SShA [Specifics of the USE in the Context of Educational Testing in the US]. *Voprosy obrazovaniya*, no 2, pp. 234–248.